令和元年度 修士論文

音楽 TV 番組における特定ダンスシーンの符号化検索

電気通信大学大学院 情報理工学研究科
情報・ネットワーク工学専攻
電子情報学プログラム

学籍番号 : 1731172
ZHANG XUEQI

指導教員
森田 啓義　教授
笠井 裕之　教授

令和元年9月20日

目 次

第1章　はじめに

1.1　本研究の背景

近年，TVチャンネル数の増加に伴い，ユーザーが視聴できるコンテンツも益々多様化になってきており，膨大なコンテンツから興味のあるシーンを自動的に抽出する技術が強く望まれている.

このような興味のあるシーンを自動的に抽出する研究事例として，MPEG2データに含まれる各種圧縮符号化パラメータを特徴量として，HD（高画質）サッカー映像からのハイライトシーンを検出する研究[1]や，圧縮符号化パラメータおけるBBPというピクチャ列での「順方向予測→逆方向予測→フレーム内符号化」という変遷をしているマクロブロックによる移動体領域の推定を用いた投球動作の判定と動きベクトルによるカメラアングルの判定を利用して，SD（標準画質）野球映像から投球シーンの判定をする研究[2]がある．しかし，これらの研究では，検索対象は検出したいシーンが出る際に，画面上に特定のパータンがあるスポーツ映像である．例えば，サッカー映像におけるゴール付近のシーンではペナルティエリアのラインが出るというパータンや，野球映像の投球シーンでは捕手と打者の身体の大部分は投球シーンの画面において画面の右上に位置し，投手は画面の左半分に写り，体の大部分はその下半分に写るというパータンがある.

本研究では，様々なコンテンツの中から，特に音楽ライブ番組において，歌手が歌いながら行う特定のダンスシーンの自動抽出に着目した．ダンスシーンは同一グループによる同じ振り付けであっても，番組ごとに様々なカメラアングルで撮影されることが多く，ファンにとってはどれも見逃したくないだけでなく，従来の顔検出などのビデオ処理技法を適用してもダンスシーンを特徴づけることは困難である.

1.2 本研究の目的

本研究では，圧縮動画の符号化パラメータを特徴量とし，復号化をせずにより素早くHD音楽ライブ番組からダンスシーンの検出と，検出されたダンスシーンから動画検索キーによる特定ダンスシーンの検索を目指す.

そこで本研究では，平滑化z-スコアアルゴリズムと，BSpline曲線を組み合わせることによって，音楽ライブ番組におけるダンスシーンの検出手法を提案する．さらに，文献[3]の提案手法である圧縮符号化パラメータの16×8画素等のマクロブロックの出現箇所と，動きベクトルの情報の利用により，動作領域におけるそれらの出現パターンが特徴量として特定動作を抽出する手法に基づいて，既存のダンスシーンの先頭の100～500枚のフレームからなる動画クリップ（3秒～16秒の動画）を検索キーとすることによって，音楽TV番組における特定ダンスシーンの検索を行う.

1.3 本論文の構成

本論文の構成は以下の通りである．まず2章では，符号化パラメータを利用した映像解析と，音楽番組のシーン解析に関する先行研究を紹介する．3章では，符号化パラメータと音楽TV番組における特定ダンスシーンの検索を行うシステムの概要について述べる．4章では，従来手法と提案手法によるダンスシーンの検出アルゴリズム，実験，結果，考察，比較，提案手法に関する改善方法および検出されたダンスシーンの始点と終点の精度について述べる．5章では，提案手法による特定ダンスシーンの検索手順，実験，結果および考察について述べる．最後に，6章では，これまでの結果と今後の課題についてまとめる．

第2章　関連研究

本章ではこれまでの符号化パラメータを利用した映像解析と，音楽番組のシーン解析に関する研究について紹介する．2.1節において符号化パラメータを利用した映像解析の先行研究について述べる．次に，2.2節において画像と音響特徴による音楽番組のシーン解析を行った先行研究について述べる．最後，2.3節において動画クリップを検索キーとする圧縮動画検索を行った先行研究について述べる．

2.1　符号化パラメータを利用した映像解析

文献[4]ではMPEG-2の動画において，符号化パラメータの中でも16×8画素のマクロブロック（以下16×8MB）が移動体の境界領域に集中して発生することや，画面の動きの激しい部分に発生することを示し，それをを利用して移動体の検知と追跡を行っている．図2.1に移動体を写るシーンにおいて16×8MBを表示した画面を示す.

図 2.1: 移動体を写るシーンにおける16×8MB（緑色で塗られたところ）の分布，文献[4]の図3.1から引用

また，文献 [1] では，16 × 8MB 数を用いて，MPEG2 形式の HD サッカー映像における Up/Long ショットの判別を行っている．Long ショットでは，選手の姿が画面サイズに比べて小さく，動きほとんどがカメラワークに起因しているため，16 × 8MB の発生数が少なくなる傾向にある．一方，Up ショットでは，選手が大きく映っているため，Long ショットと比較すると動きベクトルの発生頻度は大きくなり，16 × 8MB 数も増える傾向にあることが示されている．図 2.2 に Up/Long ショットにおいて 16 × 8MB を表示した画面を示す.

図 2.2: Up(右)/Long(左) ショットにおける 16 × 8MB（赤色で塗られたところ）の分布，文献 [1] の図 6.4 から引用

本研究では，これらの先行研究結果に基づき，上述した 16 × 8MB の発生頻度，とくにその時間変動が急激に上昇，下向する時点に着目して，音楽 TV 番組におけるダンスシーンを検出する方法を提案する．

2.2 音楽番組のシーン解析

音楽番組のシーン解析に関しては，顔認識手法により出演者変化の検出と，音響信号が属するオーディオクラスの帰属確率により無音、音楽、音声、雑音の分類を用いて音楽番組における司会シーンと歌唱シーンの分割を行った研究 [5] や，顔認識手法に基づく口の動き検出手法と，歌声区間と非歌声区間の状態を行き来する隠れマルコフモデルによって楽曲を表現することで混合音中の歌声区間を推定する手法を組み合わせることによって歌唱シーン検出をする研究 [6] が報告される．

しかし，画像と音響特徴のみでは，出演者のからだ全体の動きを把握できないため，検出された歌唱シーンにダンスパフォーマンスが含まれているかどうかまでを判断することはできない．

2.3 動画クリップを検索キーとする圧縮動画検索

動画クリップを検索キーとする圧縮動画検索に関しては，文献[3]では，H.264/AVC形式の動画から符号化パラメータである可変マクロブロック(16×8MB，8×16MB，8×8MB)と動きベクトルの情報によって動作領域を特定し，その領域を3Dモデルソフトで3次元空間上にフレームごとにオブジェクトを設置して3Dモデルとして表現し，その重心をフレームごとに繋ぐことで動作の時空間的な変移を3D重心曲線として表現した．検索キー動画と検索対象それぞれから得られた3D重心曲線をマッチングすることよって動作の検索を行っている．

図2.3，図2.4に，H.264/AVC形式の動作動画から作成した3Dモデルと3D重心曲線の例を示す．

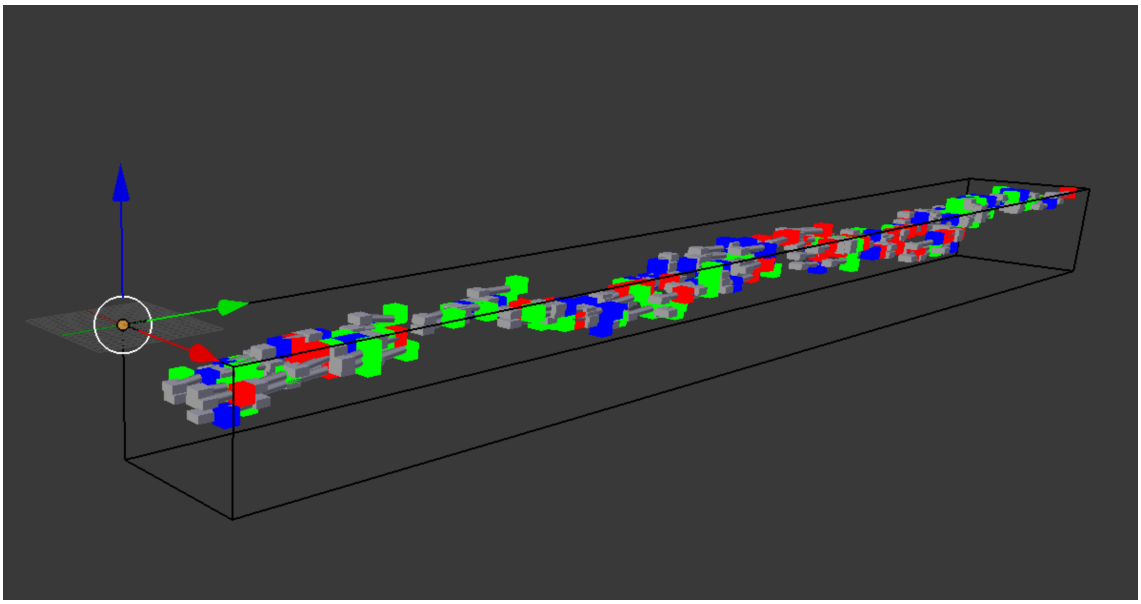

図 2.3: 3Dモデルの例，文献[3]の図3.4から引用

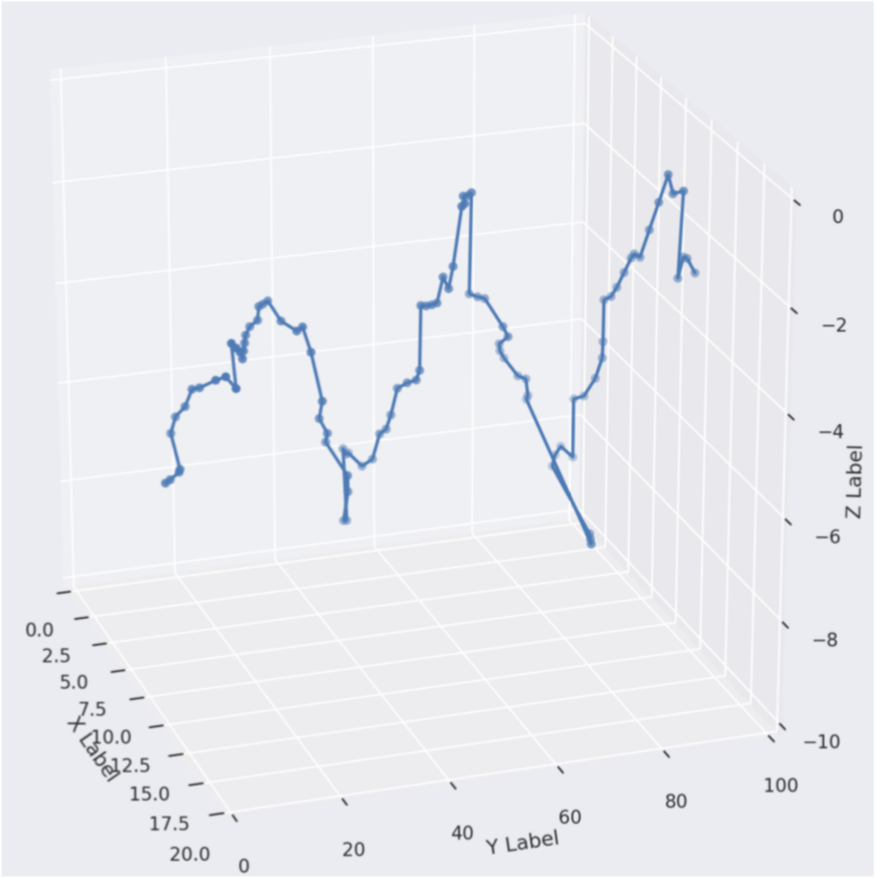

図 2.4: 3D 重心曲線の例，文献 [3] の図 3.9 から引用

図 2.3 において，立方体は，青色が 16×8，赤色が 8×16，緑色が 8×8 画素の MB を表す．また，灰色の円柱は動きベクトルを表し，灰色の立方体は動きベクトルの始点座標を表す．動きベクトルの終点座標は全て，いずれかの MB の中心座標に対応する．赤矢印が x 座標，緑矢印が y 座標，青矢印が z 座標にそれぞれに対応する．

本研究では，特定ダンスシーンの検索について，出演者が真正面から映り，カメラワークの使用が少ない傾向にあることに着目して，既存のダンスシーンの先頭 6 秒 ~ 15 秒の動画を検索キーとし，文献 [3] の提案手法に基づき，特定ダンスシーンの検索を行う．

第3章　提案システム

本章では，本研究で使用した符号化パラメータと音楽TV番組における特定ダンスシーンの検索を行うシステムを構成する概要について述べる．

3.1　符号化パラメータ

MPEG -2（地上デジタル放送の標準規格である符号化法）では1秒間に約30枚のフレーム（ピクチャとも言う）がI、P、Bという3種類のピクチャに分類され，それぞれのピクチャに定められた予測符号化法によって高効率なデータ圧縮が実現されている．表3.1にI、P、Bピクチャにおいて使用している予測符号化法（“○”マーク）を示す． Intra符号化は他のピクチャを参考せず，ピクチャ内の情報だけを利用した符号化である．Intra符号化に加えて，Pピクチャでは，直前のI/ Pピクチャを利用した順方向予測符号化を行い，Bピクチャでは,直前と直後のI/ Pピクチャを利用した順方向、逆方向、双方向予測符号化を行っている．図3.1，図3.2，図3.3に，圧縮動画におけるI、P、Bの例を示す.

	I	P	B
Intra符号化	○	○	○
順方向予測符号化		○	○
逆方向予測符号化			○
双方向予測符号化			○

表3.1: I、P、Bピクチャにおいて使用している予測符号化法

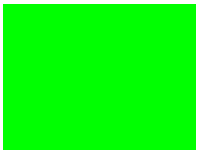

図 3.1: I

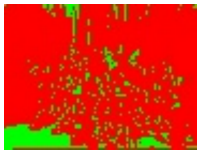

図 3.2: P

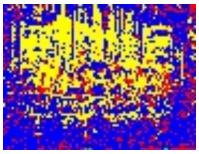

図 3.3: B

図 3.1，図 3.2，図 3.3 において，緑色で塗られたところは Intra 符号化を使用している部分であり，赤色で塗られたところは順方向予測符号化を使用している部分であり，青色で塗られたは逆方向予測符号化を使用している部分であり，黄色で塗られたは双方向予測符号化を使用している部分である．

MEPG-2 で定められた時間的に連なったピクチャを一定枚数まとめる単位である GOP は，通常は 15 枚のピクチャからなり，その中に B ピクチャは 10 枚含まれている．形式は IBBPBBPBBPBBPBB のようになっている．

また，各ピクチャは, マクロブロック (以下 MB) と呼ばれる単位に分割され，MB ごとに符号化処理が行われる．さらに，MB のサイズには 16×16 と 16×8 画素の 2 種類があり，それらの中から符号化効率が尤も良くなるように決定される．

MPEG-2 ではもう一つ重要な符号化パラメータとして, 動きベクトルが用いられている. 動きベクトルは，ある基準となるピクチャ（現画像）からもう一方のピクチャ（予測画像）間で特定の物体や領域移動した動きをベクトルで表現した情報であり, 上述した予測符号化で圧縮効果を高めるため用いられている. 図 3.4 に動きベクトルを示す.

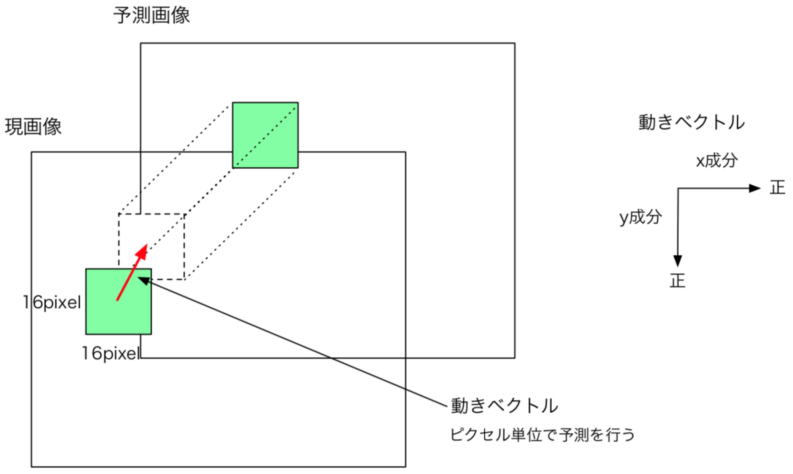

図 3.4: 動きベクトル，文献 [7] の図 3.2 から引用

3.2 提案システムのフローチャート

音楽TV番組における特定ダンスシーンの検索システム全体のフローチャットは図3.5のようになっている．提案システムは主に2つの部分からなる．第1部分（図5に赤枠で囲まれるところ）では，音楽TV番組映像からダンスシーンの検出を行う．第2部分（図5に青枠で囲まれるところ）では，音楽TV番組映像から（複数のダンスシーン映像から）検索キーと同じダンスシーン（特定ダンスシーン）の検索を行う．第1部分では，圧縮符号化パラメータを特徴量とし，平滑化z-スコアアルゴリズムと，BSpline曲線を組み合わせることによって音楽TV番組における特定ダンスシーンの検索手法を提案する．具体的な手法および実験は4章で述べる．第2部分では，文献[3]の提案手法に基づいて複数のダンスシーン映像から特定ダンスシーンを検索した．具体的な手順および実験は5章で述べる．

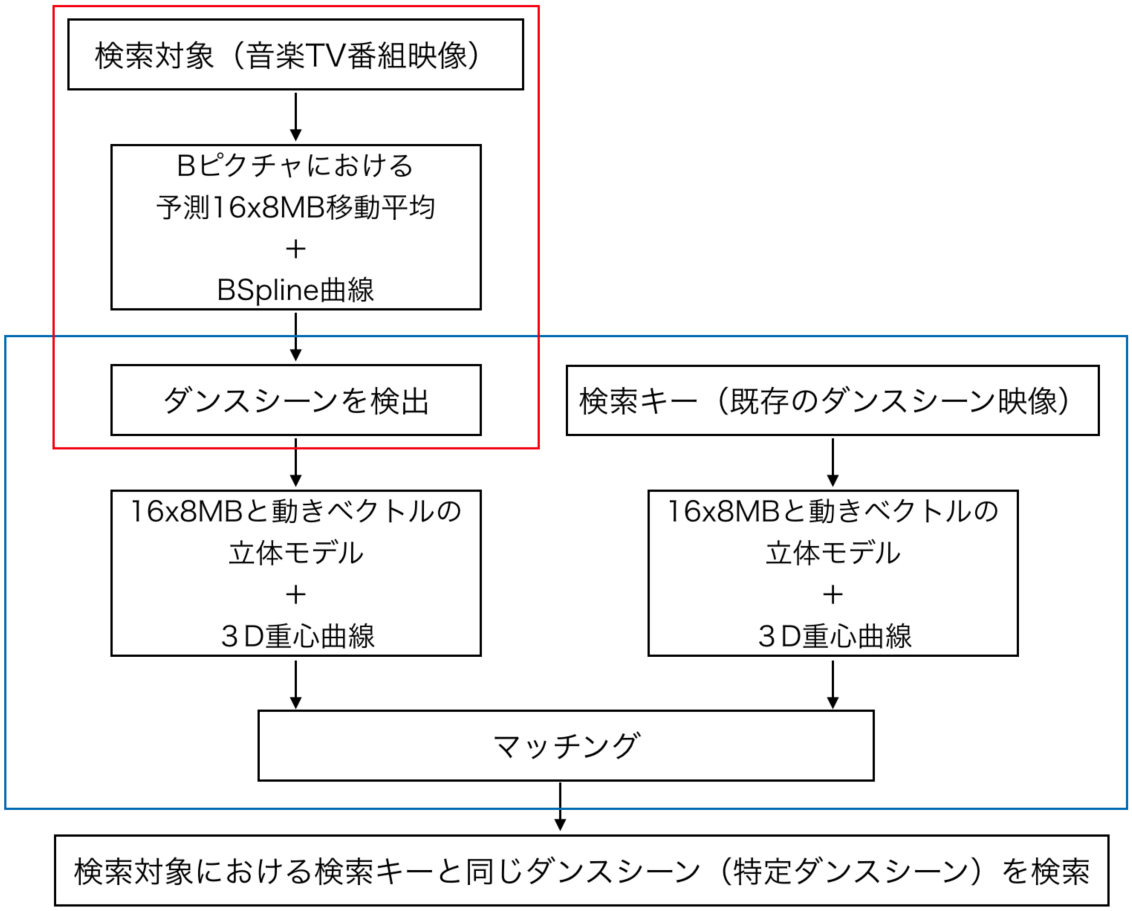

図 3.5: 提案システムのフローチャート

第4章　ダンスシーンの検出

本章では，提案システム第1部分の実験について述べる．まず，4.1節で実験環境について述べる．4.2節で圧縮動画におけるダンスシーンの特徴について述べる．4.3節で従来手法平滑化z-スコアアルゴリズム，実験，結果および考察について述べる．4.4節で提案手法アルゴリズム，実験，結果，考察，従来手法との比較について述べる．4.5節で提案手法の改善について述べる

4.1　実験環境

本節では，実験を行った動作環境について述べる．本研究の全ての実験には，Mac Book Pro（OS : Mac OS 10.13.6，CPU周波数 : 2.8 GHz，メモリ : 16 GB）を用い，圧縮符号化パラメータの取得はMplayer[8]のffmpegライブラリ[9]を用いて行った．ダンスシーンの検出および特定ダンスシーンの検索はPython 3.7.2を用いて作成したプログラム上で解析処理を行った．

4.2　圧縮動画におけるダンスシーンの特徴

2.1節で述べた16×8MBのプロパティに基づいて，本研究では，Bピクチャにおいて順方向、逆方向、双方向予測符号化を行っている16×8MB（以下予測16×8MB）の数に着目した．

まず予備実験として，9本の音楽番組映像（再生時間は9分~15分）を使用し, GOPごとに予測16×8MBの数を調べた．

実験に用いる音楽番組映像9本のうち，3本には少なくとも一つ以上のダンスシーンが含まれていて，残りの6本はダンスシーンは含まれていない．図4.1(a)はダンスシーンを含むある1本の映像の結果であり，一方，ダンスシーンのない映像の結果の一例を図4.1(b)に示す. 両図とも，横軸はGOP番号を表し, 縦軸は予測16×8MBの総数を表す. 図4.1(a)においてはGOP番号153~486の区間 (黒枠で囲まれる部分) はダンスシーンである.

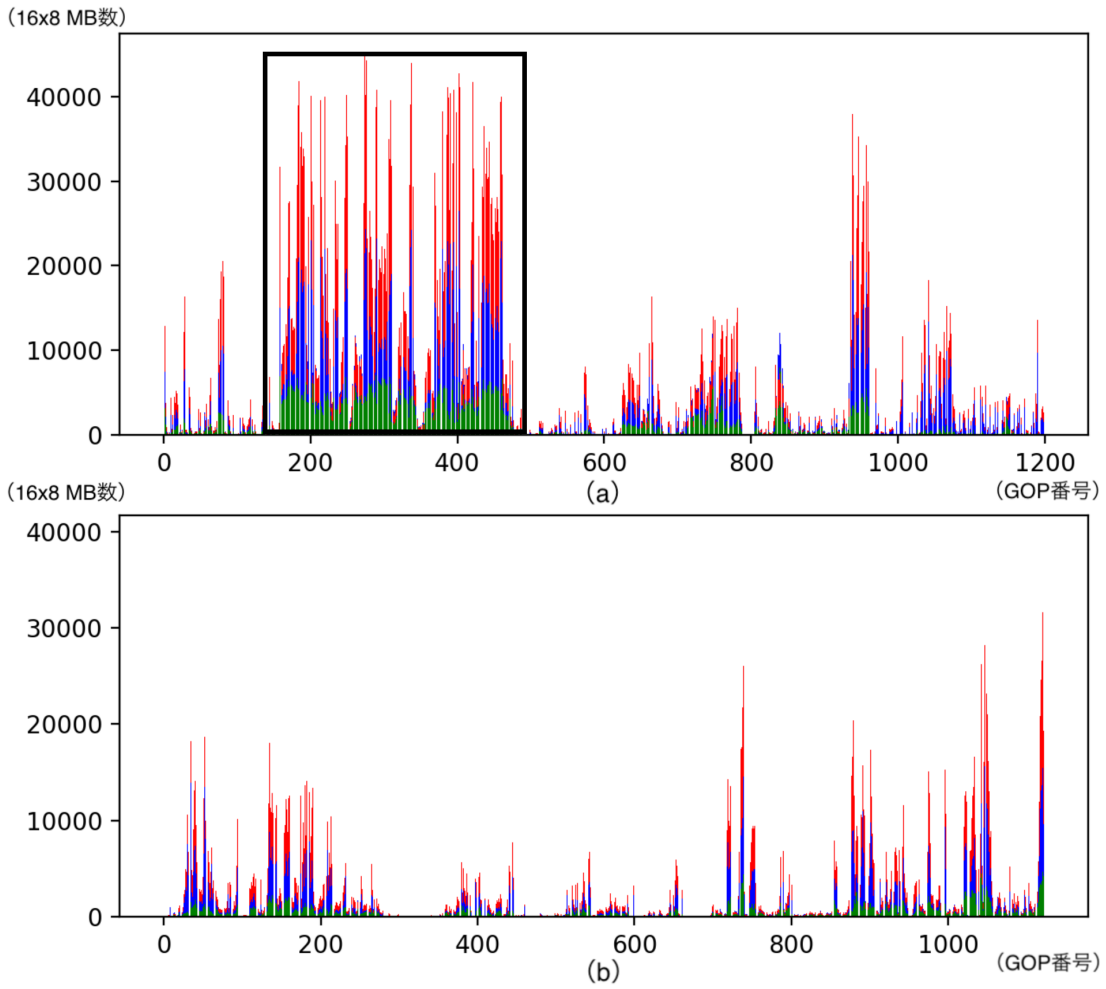

図 4.1: GOP ごとに予測 16×8MB の数．積み上げ棒について, 緑の部分は双方向予測符号化 16×8MB であり, 青の部分は逆方向予測符号化 16×8MB であり, 赤の部分は順方向予測符号化 16×8MB である.

9 本の映像を用いて得た知見をまとめると, ダンスシーンでは，予測 16×8MB が発生しやすく、20000 以上の数の 16×8MB が発生した GOP が多いことと，ダンスシーンの始まりと終わりの GOP 番号において，急激な 16×8MB 数の変化が発生することである．

4.3 従来手法

離散信号のピーク検出については，過去の一定の長さの信号の平均と標準偏差に基づいて，現時点のピークを検出する平滑化 z-スコアアルゴリズムが報告されている [10][11]. 同アルゴリズムでは，移動平均を用い，過去の信号を平滑化するので，次のピークを検出する際に，現時点のピークからの影響は程よく減ることが可能であると考えられる．

4.3.1 平滑化 z-スコアアルゴリズム

平滑化 z-スコアアルゴリズムを [Algorithm 1] に示す 10 ステップからなる．

Algorithm 1 平滑化z-スコアアルゴリズム

入力: データ列$\boldsymbol{Y} = Y_1 Y_2 Y_3 \ldots Y_n$

出力: 判定結果$\boldsymbol{O} = O_1 O_2 O_3 \ldots O_n$

- **内部変数**ℓ：スライド窓の長さ（$\ell \geq 2$）
  w：重み係数
  τ：閾値

1. **for** i=1, 2, 3$\cdots$, n, ..., $n+l-1$ **do**
2. $\mathrm{S}_i = wY_i + (1-w)\,\mathrm{S}_{i-1}$ (1)
   を求める．ここで，$\mathrm{S}_0 = 0$.
3. 移動平均$\overline{S_i}$を求める．

$$\overline{S}_i = \begin{cases} \frac{1}{i}\sum_{k=1}^{i} \mathrm{S}_k & 1 \leq i \leq l-1 \\ \frac{1}{l}\sum_{k=i-l+1}^{i} \mathrm{S}_k & l \leq i \leq n \\ \frac{1}{n-i+l}\sum_{k=i-l+1}^{n} \mathrm{S}_k & n+1 \leq i \leq n+l-1 \end{cases} \quad (2)$$

4. 移動標準偏差σ_iを求める．

$$\sigma_i = \begin{cases} 0 & 1 \\ \sqrt{\frac{1}{i-1}\sum_{k=1}^{i} \left(\mathrm{S}_k - \overline{S}_i\right)^2} & 2 \leq i \leq l-1 \\ \sqrt{\frac{1}{l-1}\sum_{k=i-l+1}^{i} \left(\mathrm{S}_k - \overline{S}_i\right)^2} & l \leq i \leq n \\ \sqrt{\frac{1}{n-i+l-1}\sum_{k=i-l+1}^{n} \left(\mathrm{S}_k - \overline{S}_i\right)^2} & n+1 \leq i \leq n+l-1 \end{cases} \quad (3)$$

5. **end for**
6. $z_1 = 0, z_2 = 0, O_1 = 0, O_2 = 0$とおく
7. **for** i=3, 4$\cdots$$n$ **do**
8. $\overline{S_{i-1}}$とσ_{i-1}を用いて，Y_iを正規化する．

$$z_i = \frac{Y_i - \overline{S}_{i-1}}{\sigma_{i-1}} \quad (4)$$

9. 判定

$$O_i = \begin{cases} 1 & \text{if } |z_i| \geq \tau \\ 0 & \text{if } |z_i| < \tau \end{cases} \quad (5)$$

10. **end for**

4.3.2 平滑化z-スコアを用いたダンスシーン検出の実験および考察

平滑化z-スコアアルゴリズをダンスシーン検出に適用するにあたって，入力データや各パラメータを以下のように定めた：

入力データ列は，GOP単位でまとめたBピクチャにおける予測16×8MBの総数$\boldsymbol{Y} = Y_1Y_2Y_3 \ldots Y_n$であり，nはGOP数である．さらに，$\ell$=30，$w$=0.5とおく．特に，閾値$\tau$の決め方は，データに依存するので，本研究では式(4)で得たデータ列z_iの累積分布関数をcdf(t)とおき，0.969≤ cdf(t) ≤0.981を満たすtの区間の中点を閾値τと定めた．再生時間は9分∼15分である10本のダンスシーンありの音楽番組映像を用いて実験を行った．全ての映像に少なくとも1つ以上のダンスシーンが含まれている．実験結果は表4.1に示す．表4.1において，ダンスシーンの正解（始点-終点）および結果（検出された始点-検出された終点）の単位はGOP番号である．

映像No.	GOP数	ダンスシーンの正解	結果	誤差（Δ始点/Δ終点）
1	1200	153-486	159-462	-6/-24
2	1663	398-758	377-762	+21/+4
		122-424	73-356	＋49/-68
3	1928	654-847	x	x
		848-1042	x	x
4	1829	1232-1554	1250-1623	-18/+69
5	1797	326-646	x	x
6	1771	1396-1768	x	x
7	1751	16-440	1346-1632	＊
8	1879	324-636	342-652	-18/+16
9	1657	336-676	321-600	+15/-76
10	1471	258-682	279-762	-21/+80

表4.1: 平滑化z-スコアを用いたダンスシーン検出の結果．“x”はダンスシーンを検出できないこと（プログラムで出力はNULLであること，以下未検出）を表す．“＊”は結果の誤差がかなり大きいこと（以下検出外れ）を表す．

誤差における“Δ始点”は検出された結果の始点と正解のダンスシーンの始点の差であり，“＋”は正解のダンスシーンの始点より前のGOPで検出されたことを表し，“－”は正解のダンスシーンの始点より後のGOP

で検出されたことを表す. 誤差における“Δ終点”は検出された結果の終点と正解のダンスシーンの終点の差であり,“＋”は正解のダンスシーンの終点より後のGOPで検出されたことを表し,“－”は正解のダンスシーンの終点より前のGOPで検出されたことを表す.

表4.1より，平滑化z-スコアアルゴリズムを用いたダンスシーンの検出手法では，映像3、5、6に対して未検出が発生し，映像7に対して検出外れが発生したことが分かった.

4.4 提案手法

平滑化z-スコアアルゴリズムを用いた実験の結果を踏まえて，閾値のτの決め方による検出精度への影響を減らすため，平滑化z-スコアアルゴリズムの後半で行っているY_iの正規化と，閾値を用いた判別を行わないで，移動平均の部分(式(1) ∼ (2))と，新たにBSpline曲線[12][13][14]による平滑化を組み合わせる手法を提案する.

4.4.1 BSpline曲線の定義

BSpline曲線は，BSpline係数とノットより定義される区分的多項式で表された曲線である。具体的な定義は[定義1]に示す. 点列$(i,\ Y_i),(i=1,2,3,...,n)$を補間するBSpline曲線$S(x)$を求めるには，与えられた点列の$x$座標定義域$[1,n]$から，等間隔に選んだ一定数（以下近似点数$N_B$，$N_B \in \mathcal{Z}^+$）の$x$座標を取得し，各$x$座標を用いて，式（6）によって$S(x)$を求める．最後に，それらの$(x,S(x))$が座標となる一定数の点を繋げて補間した曲線がBSpline曲線である.

定義 1 BSpline 曲線 $S(x)$

基本式：$\mathrm{S}(x) = \sum_{j=1}^{n} c_j B_{j,k;t}(x)$ (6)

内部変数：

（i）BSpline 多項式の次数：k

（ii）BSpline 係数：$\mathrm{c} = c_1 c_2 ... c_n, n \geq k+1$,

c は与えられた点列の y 座標値により求める.

$$c_j = \begin{cases} Y_1 & j = 1 \\ (1-\lambda)Y_{j-1} + \lambda Y_j & \lambda \in \mathcal{R}, j = 2, 3, ..., n-1 \\ Y_n & j = n \end{cases} \quad (7)$$

（iii）ノット：$\mathrm{t} = t_1 t_2 ... t_u, u = n+k+1$,

t は与えられた点列の x 座標の閉区間により求める.

$$\text{k=3 の場合，} t_i = \begin{cases} 0 & i = 1, 2..., k+1 \\ i-k & i = k+2, k+3, ..., n \\ n-1 & i = n+1, n = 2, ..., u \end{cases} \quad (8)$$

NOTES：

$$B_{i,0}(x) = \begin{cases} 1 & \text{if } t_i \leq x \leq t_{i+1} \\ 0 & otherwise \end{cases} \quad (9)$$

$$B_{i,k}(x) = \frac{x - t_i}{t_{i+k} - t_i} B_{i,k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} B_{i+1,k-1}(x) \quad (10)$$

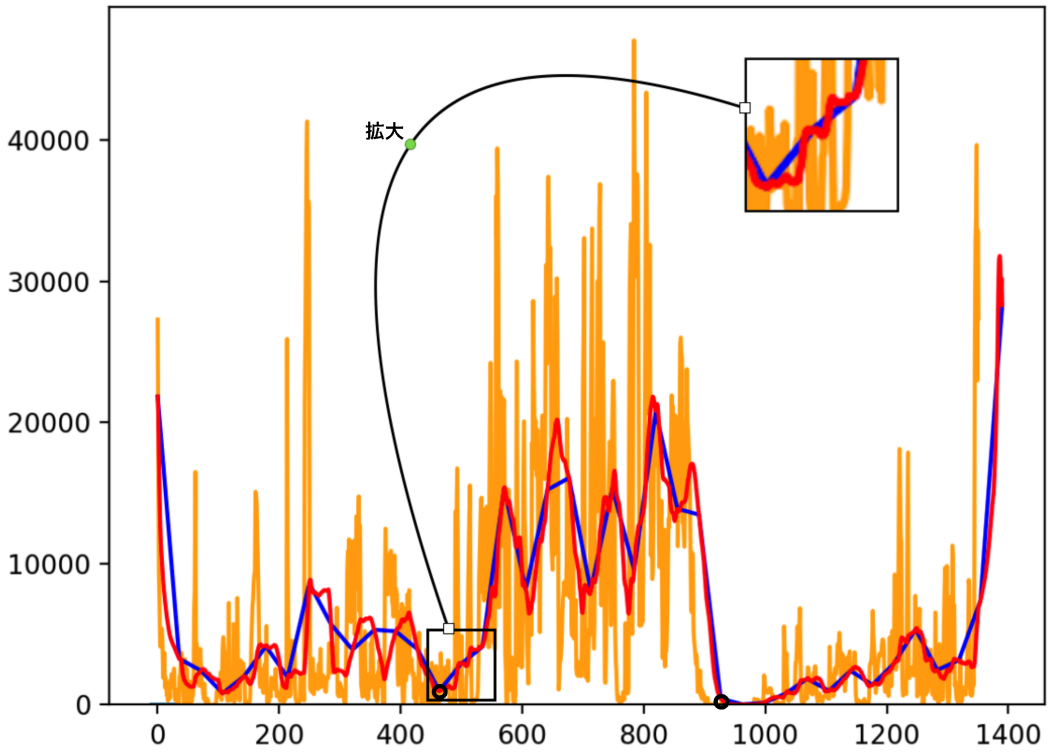

図 4.2: ダンスシーンありの映像に 16×8MB 数の移動平均折れ線と BSpline 曲線．予測 16×8MB の総数はオレンジ線であり，16×8MB 総数の移動平均折れ線は赤い線であり（スライド窓の長さ $\ell = 30$），移動平均折れ線の BSpline 曲線は青い線である．k =3, BSpline 係数 c の数＝移動平均の長さ（1381），ノット t の数=1385，近似点数（N_B）＝ 50.

予備実験の観察からダンスシーンの始まりと終わりのGOP番号において，急激な16×8MB数の変化が発生することが分かっているので，図4.2において，黒い丸でマークする極小値に対するx座標（極小点と呼ぶ）がダンスシーンの始点と終点であると考えられる．元の点列に比べ，$S(x)$上では極小点の数が大幅に減少した．

4.4.2 BSpline曲線によるダンスシーン検出のアルゴリズム

BSpline曲線上におけるダンスシーンの始点と終点である可能性が高い極小点は，それらを検出するために，以下を用いる．

• k=3

• $S(x)$の定義域上で等間隔にとったx座標の集合：

$\xi = \{\xi_1\ ,\ \xi_2\ ,\ \xi_3\ ,\ ...,\xi_n\}$,

$\xi_i = \frac{GOP\ 数+l-2}{n-1} \times (i-1)+1 \quad (1 \leq i \leq n, n = N_B)$.

• 極小点の集合：$\mathcal{S}_{\min} = \{\underline{x}_1\ ,\ \underline{x}_2\ ,\ \underline{x}_3\ ,... \underline{x}_k\}$（$0 \leq k \leq n$），$\mathcal{S}_{\min} \subset \xi$.

•BSpline曲線上の極大値に対するx座標を極大点と呼ぶ．極大点の集合：$\mathcal{S}_{\max} = \{\overline{x}_1\ ,\ \overline{x}_2\ ,\ \overline{x}_3\ ,... \overline{x}_m\}$（$0 \leq m \leq n$），$\mathcal{S}_{\max} \subset \xi$.

• 極点（極小点と極大点）の集合：$\mathcal{M} = \mathcal{S}_{\min} \cup \mathcal{S}_{\max} = \{x_1\ ,\ x_2\ ,\ x_3\ ,... x_{k+m}\}$ $(x_i < x_{i+1}, 0 \leq k+m \leq n)$，$\mathcal{M} \subset \xi$.

BSpline曲線によるダンスシーン検出のアルゴリズムは[Algorithm 2]に示す．検出手順においては，$\mu_1 = 7000, \mu_2 = 4500$とおく．

Algorithm 2 提案アルゴリズム

入力: ダンスシーンがある音楽番組映像における予測16×8MB移動平均のBSpline曲線$\mathrm{S}(x)$

出力: 始点集合$\mathcal{X}_i^s$, 終点集合$\mathcal{X}_i^e$

1: **for** i=$1, 2, 3 \cdots k+m$ **do**

2: もし，$x_i \in \mathcal{S}_{\min}$，$x_{i+1}\, x_{i+2} \in \mathcal{M}$であり，さらに，2つの条件
1) $B(x_i) \leq \mu_1$
2) $\min\{B(x_{i+1}) - B(x_i), B(x_{i+2}) - B(x_i)\} \geq \mu_2$
が全て真であるならば，x_iを$\mathcal{X}_i^s$に代入する．

3: **end for**

4: $\mathcal{X}_i^s$を$\{x_1^s, x_2^s, ..., x_l^s\}$とおき，
for i=$1, 2, 3 \cdots l$ **do**

5: **for** j=$1, 2, 3 \cdots k+m$ **do**

6: もし，$x_j \in \mathcal{S}_{\min}$，$x_{j-1}\, x_{j-2} \in \xi$であり，さらに，2つの条件
1) $250 \leq x_j - x_i^s \leq 510$
2) $\min\{B(x_{j-1}), B(x_{j-2})\} > B(x_j)$
が全て真であり，x_jの数は1であるならば，x_jを$\mathcal{X}_i^e$に代入する．もし，x_jの数は1以上なら，それらの平均を$\mathcal{X}_i^e$に代入する．

7: **end for**

8: **end for**

9: $\mathcal{X}_i^s$と$\mathcal{X}_i^e$の中から，GOP区間の端点に対応したペアについて，16000以上の予測16×8MB数をもつGOPの数が60個以上あれば，そのペアをダンスシーンの端点でとして出力する．

4.4.3 提案手法を用いたダンスシーン検出の実験および考察

4.3節で使用した映像データを利用して提案手法を用いて予備実験を行った，結果は表4.2に示す.

映像 No.	1	2	3	4	5	6	7	8	9	10
Δ 始点	+28	+53	+3, +16, x	+20	+66	+75	-56	+13	-7	+16
Δ 終点	-40	+12	+55, +75, x	+37	-26	-67	+31	+103	+28	-25

表 4.2: BSpline を用いたダンスシーン検出の予備実験結果.ℓ=30, w=0.8, N_B=50.

提案手法を用いた予備実験の結果を観察すると，映像3に対して未検出が発生した．未検出の映像本数については，平滑化z-スコアアルゴリズムを用いた実験の結果（以下従来手法）より少なくなっている．

さらに，検出外れの映像はなかった．提案手法を検証するため、ほかの10本のダンスシーンありの音楽番組映像（映像の情報は表4.3に示す）を用いて本実験を行った．全ての映像に 1 つのダンスシーンが含まれている．wの値は0.5 以上なら，大きな変化が出ないため，wを0.8に設定した．本実験の結果を表4.4に掲げる．

映像 No.	1	2	3	4	5	6	7	8	9	10
GOP 数	1352	1145	1795	1797	1269	1797	1729	1375	1557	1258
ダンスシーンの区間	516 | 892	661 | 1042	1111 | 1455	296 | 614	935 | 1256	107 | 425	876 | 1384	51 | 420	276 | 583	21 | 474

表 4.3: 提案手法を用いた本実験用の映像の情報

映像No.	$\ell = 30$、$N_B = 50$	$\ell = 40$、$N_B = 50$	$\ell = 30$、$N_B = 40$
1	+66/-57	+34/+72	+53/+70
2	+39/-49	+10/+12	+24/-57
3	-79/+144	-11/+78	△ -17/+49
4	-1/+130	-3/+134	+155/-97
5	△ +36/+14	△ +29/+24	△ +31/+17
6	+107/-16	+107/-28 △	+107/-96 △
7	-20/-40	+11/-176	+61/-184
8	-6/+38	-6/+12	+15/+14
9	-15/+128	△ +146/-95	△ +154/-93
10	+21/-81	+21/-78	+21/-76

表 4.4: BSplineを用いたダンスシーン検出の本実験結果.“△”はダンスシーンでないシーンをダンスシーンとして検出されたこと（以下誤検出）を表す.

ダンスシーンを見逃さないため，誤差については，“－”より“＋”の方が良いと考えられる．本実験の結果を見ると，$\ell = 30$，$N_B = 40$と，$\ell = 40$，$N_B = 50$の実験において，映像5、6、9に対して誤検出が発生した. $\ell = 30$，$N_B = 50$の方には，映像5に対して誤検出が発生した．

誤検出を除いてそれぞれの誤差の平均は表4.5に示す.

	$\ell = 30$、$N_B = 50$	$\ell = 40$、$N_B = 50$	$\ell = 30$、$N_B = 40$
始点	+14.8	+33.8	+60.4
終点	+21.1	-4.5	-45.3

表 4.5: BSplineを用いたダンスシーン検出の本実験誤差の平均

誤検出を考慮しない場合，$\ell = 30$，$N_B = 50$の方は，始点誤差の平均は一番小さく，終点誤差は正数である．ℓ=30，w=0.8，N_B=50の方がより良い結果を得られることが分かった．

誤検出を考慮する場合には，各手法を用いた検出実験の結果の*Precision*，*Recall*とF値により評価する．それぞれの定義は以下になる．

- TP：正しく検出できたダンスシーンのGOP区間
- FP：ダンスシーンでないが，検出してしまったGOP区間
- FN：ダンスシーンであるが，検出できなかったダンスシーンのGOP区間
- $Precision\,(P) = \frac{TP}{TP+FP}$，$Recall\,(R) = \frac{TP}{TP+FN}$
- $F = 2 \times \frac{P \times R}{P+R}$

各手法において，各映像の*Precision*，*Recall*とF値を求め，10本の映像の*Precision*，*Recall*とF値の平均を取った．それぞれの平均は表4.6に示す.

	$\ell = 30$、$N_B = 50$	$\ell = 40$、$N_B = 50$	$\ell = 30$、$N_B = 40$
Precision の平均	78.2%	73.3%	66.8%
Recall の平均	90.8%	90.3%	83.7%
F 値の平均	0.83	0.78	0.72

表 4.6: BSpline を用いたダンスシーン検出の本実験結果の *Precision*，*Recall* と F 値の平均

$\ell = 30$，$N_B = 50$ の方に対して，$\ell = 40$，$N_B = 50$ の結果を比較すると，*Precision* は約 4.9%，*Recall* は約 0.5%，F 値は約 0.05 高い結果が得られた．$\ell = 30$，$N_B = 40$ の結果を比較すると，*Precision* は約 11.4%，*Recall* は約 7.1%，F 値は約 0.11 高い結果が得られた．誤検出を考慮する場合にも，ℓ=30，w=0.8，N_B=50 の方はより優れる結果を得られることが分かった．

4.5 提案手法の改善について

4.5.1 改善方法

4.4 節で，式（2）によって得た移動平均の全ての点列を用いて生成した BSpline 曲線上における等間隔に選んだ一定数の出力点を利用してダンスシーンの検出実験を行った．本節では，提案手法の改善を検討するため，移動平均の部分点列（以下 P）を用いて生成した BSpline 曲線上における等間隔に選んだ一定数の出力点（N_B）を利用して 10 回のダンスシーンの検出実験を行った．実験用の映像は 4.4.3 節で使用した映像である．映像の情報は表 4.3 に示す．10 回それぞれの手法は以下の方式 1~10 になっている．

方式 1： 移動平均の 50 個の点列（$P = 50$）を用いて生成した BSpline 曲線上における等間隔に選んだ 50 個の出力点（$N_B = 50$）を利用してダンスシーンの検出実験を行った

方式 2： $P = 50$，$N_B = 300$

方式 3： $P = 50$，$N_B = 1000$

方式 4： $P = 50$，$N_B = 10000$

方式 5： $P = 300$，$N_B = 50$

方式6： $P = 300$，$N_B = 300$

方式7： $P = 300$，$N_B = 1000$

方式8： $P = 300$，$N_B = 10000$

方式9： $P = 700$，$N_B = 50$

方式10： $P = 1000$，$N_B = 50$

部分点列の選び方は以下になっている．ここで，移動平均の長さをl_mとおく．

- 50個の点列$P_i, i = 1, ..., 50$：

$$P_i = \begin{cases} 1 + int(\frac{l-2}{48}) \times (i-1) & i = 1, 2..., 49 \\ \\ l_m & i = 50 \end{cases}$$

- 300個の点列$P_i, i = 1, ..., 300$：

$$P_i = \begin{cases} 1 + int(\frac{l_m-1}{299}) \times (i-1) & i = 1, 2..., 249 \\ \\ int(\frac{\frac{l_m-1}{299}+1}{50}) \times (i-250) + (1 + int(\frac{l_m-1}{299}) \times 249) & i = 250, 252, ..., 299 \\ \\ l_m & i = 300 \end{cases}$$

• 700 個の点列 $P_i, i = 1, ..., 700$：

$$P_i = \begin{cases} 1 + int(\frac{l_m - 1}{699}) \times (i-1) & i = 1, 2..., 599 \\ int(\frac{\frac{l_m-1}{699}+1}{100}) \times (i-600) + (1 + int(\frac{l_m-1}{699}) \times 599) & i = 600, 601, ..., 699 \\ l_m & i = 700 \end{cases}$$

• 1000 個の点列 $P_i, i = 1, ..., 1000$：

$$P_i = \begin{cases} 1 + int(\frac{l_m - 1}{999}) \times (i-1) & i = 1, 2..., 799 \\ int(\frac{\frac{l_m-1}{999}+1}{200}) \times (i-800) + (1 + int(\frac{l_m-1}{999}) \times 799) & i = 800, 601, ..., 999 \\ l_m & i = 1000 \end{cases}$$

方式1~5の結果は表4.7に示し，方式6~18の結果は表4.8に示す.

映像No.	方式1	方式2	方式3	方式4	方式5
1	+66/-85	-5/-8	-4/-8	-4/-8	+66/-57
2	+15/-49	+6/-48	+6/-49	+6/-49	+39/+35
3	-79/+132	-71/+126	-75/+127	-74/+126	-5/+70 △
4	-1/+93	-3/+87	+1/+88	+0/+88	-1/+130
5	+406/-423	+411/-361	+410/-361	+410/-361	+433/-423
6	+107/-16	+107/-23,△	+107/-25,△	+107/-25,△	+107/-16
7	+52/-273	+7/-192	+4/-155	+5/-155	-20/-40
8	△ -6/-20 △	△ 0/+4 △	△ +1/+4 △	△ +1/+4 △	△ -6/+38 △
9	△ +18/+63	△ +6/+63	△ +7/+65	△ +7/+65	△ -15/+128
10	+21/-81	+21/-82	+21/-81	+21/-81	+21/-81

表 4.7: 方式1~5の結果

映像No.	方式6	方式7	方式8	方式9	方式10
1	-84/+129 △,△	-194/+176 △	-27/+54 △,△	+66/-57	+66/-57
2	△,-18/+36	△,+12/+36	△,+12/+37	+15/-13	+39/-133
3	-71/+105	-75/+111	-74/+111	-79/+144	-5/+70 △
4	-3/+47,△	△,+1/+41 △,△	△,+2/+41 △,△	-1/+130	-1/+130
5	△,△ +380/+327	△,△ +3/-15	△,△,△ +3/-15	△ +36/-26	-380/-357
6	△,+107/-21	+107/-30	+107/-21	+107/-16	+107/-16
7	-11/-126	-12/-150	-12/-150	-20/-40	-20/-40
8	△ 0/+34 △, △	△ 0/+7 △, △	△ -1/+9 △, △	△ -6/+38 △	△ -6/+38 △
9	△ -26/+92 △	+99/+1 △ △	+99/+4 △ △	△ -15/+128	△ -15/+128
10	+21/-83,△	+21/-84, △	+21/-70, △	+21/-81	+21/-81

表 4.8: 方式6~10の結果

表4.7，表4.8を見ると，使用した移動平均部分点列の数が等しい場合，BSpline曲線の出力点を増やすと，検出されたダンスシーン始点と終点の誤差が小さくなる傾向にある．一方，BSpline曲線の極小値の数が多くなるため，誤検出も多くなる．

誤検出の削減については，提案手法の結果に誤検出が少ないため，提案手法の結果を基準にして，各映像に対して，各方式を用いて得た結果から，提案手法の結果と最も近い結果（差分が最も小さい結果）をダンスシーン区間として出力することを提案する．各方式の誤検出を減らした結果は表4.9と表4.10に示す．

映像No.	方式1	方式2	方式3	方式4	方式5
1	+66/-85	-5/-8	-4/-8	-4/-8	+66/-57
2	+15/-49	+6/-48	+6/-49	+6/-49	+39/+35
3	-79/+132	-71/+126	-75/+127	-74/+126	-5/+70
4	-1/+93	-3/+87	+1/+88	+0/+88	-1/+130
5	+406/-423	+411/-361	+410/-361	+410/-361	+433/-423
6	+107/-16	+107/-23	+107/-25	+107/-25	+107/-16
7	+52/-273	+7/-192	+4/-155	+5/-155	-20/-40
8	-6/-20	0/+4	+1/+4	+1/+4	-6/+38
9	+18/+63	+6/+63	+7/+65	+7/+65	-15/+128
10	+21/-81	+21/-82	+21/-81	+21/-81	+21/-81

表 4.9: 誤検出を減らした方式1~5の結果

映像No.	方式6	方式7	方式8	方式9	方式10
1	-84/+129	-194/+176	-27/+54	+66/-57	+66/-57
2	-18/+36	+12/+37	+12/+37	+15/-13	+39/-133
3	-71/+105	-75/+111	-74/+111	-79/+144	-5/+70
4	-3/+47	+1/+41	+2/+41	-1/+130	-1/+130
5	+380/-327	△ +3/-15	△ +3/-15	△ +36/-26	+380/-357
6	+77/+7	+107/-30	+107/-21	+107/-16	+107/-16
7	-11/-126	-12/-150	-12/-150	-20/-40	-20/-40
8	0/+34	0/+7	-1/+9	-6/+38	-6/+38
9	-26/+92	-25/+112	+26/+105	-15/+128	-15/+128
10	+21/-83	+21/-84	+21/-70	+21/-81	+21/-81

表 4.10: 誤検出を減らした方式6~10の結果

4.5.2 改善した各方式に対する評価

各方式を評価するために，映像5に対して，多くの方式では大きな誤差が出たため，映像5を除いて，4.4節での提案手法であるℓ=30，w=0.8，N_B=50（以下提案手法）を含めて各方式と，誤検出を減らした各方式を用いた検出結果における10本映像のF値の平均を取った．結果は表4.11，表4.12に示す.

	提案手法	方式1	方式2	方式3	方式4	方式5
F値の平均	0.85	0.77	0.80	0.81	0.81	0.80
誤検出を減らした後のF値の平均	-	0.83	0.87	0.88	0.88	0.86

表 4.11: 各方式のF値の平均I

	方式6	方式7	方式8	方式9	方式10
F値の平均	0.70	0.68	0.71	0.81	0.78
誤検出を減らした後のF値の平均	0.85	0.84	0.88	0.87	0.84

表 4.12: 各方式のF値の平均II

表4.11，表4.12を見ると，方式3と方式4の検出結果がほとんど同じであるため，F値の平均も同じである．また，方式1～10の誤検出が多いため，提案手法のF値の平均はほかの方式より高い．一方，提案手法の結果を基準にして方式1～10の誤検出を減らした後のF値の平均が著しく上がった．とくに，方式２，方式3，方式4，方式5，方式8，方式9のF値の平均は提案手法より高くなった．

4.5.3 検出されたダンスシーンの始点と終点に対する評価

ユーザーにとって，検出されたダンスシーンの始点と終点に対する許容幅はそれぞれになるため，本節では，許容幅dを10 GOP，15 GOP，20 GOPに設定し，TPRおよびFPRによる提案手法，方式２，方式3，方式5，方式8，方式9を用いて検出されたの始点と終点の評価を行う．TPRおよびFPRの定義は表4.13に示す．ここで，ダンスシーンを含む映像は4.3節と4.4節の実験で使用した20本の映像である．ダンスシーンを含まない映像は再生時間は10分～15分となる20本のダンスシーンなしの音楽番組映像である．

ダンスシーンを含む映像：P　ダンスシーンを含まない映像：N Pの真である（目視で判断）ダンスシーンの始点：S Pの真である（目視で判断）ダンスシーンの終点：E 許容幅d>0		
	始点	終点
TP	Pを用いて検出された全ての始点（目視判断で，ダンスシーンであるかどうかにかかわらず）の中に，[S-d,S+d]に入っているものがある	Pを用いて検出された全ての終点（目視判断で，ダンスシーンであるかどうかにかかわらず）の中に，[E-d,E+d]に入っているものがある
FN	Pを用いて検出された全ての始点（目視判断で，ダンスシーンであるかどうかにかかわらず）が[S-d,S+d]に入ってない	Pを用いて検出された全ての終点（目視判断で，ダンスシーンであるかどうかにかかわらず）が[E-d,E+d]に入ってない
FP	Nを用いて始点を検出してしまった	Nを用いて終点を検出してしまった
TN	Nを用いて始点を検出してなかった	Nを用いて終点を検出してなかった
TPR (True Positive Rate)	TP/（TP+FN)	
FPR (False Positive Rate)	FP/（FP+TN)	

表 4.13: TPR および FPR の定義

各 d に対して，各方式を用いて検出された結果の始点と終点の TPR と FPR は表 4.11，表 4.12，表 4.13 に示す.

d=10							
		提案手法	方式2	方式3	方式5	方式8	方式9
始点	TP	3	6	7	3	5	2
	FN	17	14	13	17	15	18
	FP	1	0	0	0	0	0
	TN	19	20	20	20	20	20
	TPR	0.15	0.30	0.35	0.15	0.25	0.10
	FPR	0.05	0.00	0.00	0.00	0.00	0.00
終点	TP	0	2	2	0	2	0
	FN	20	18	18	20	18	20
	FP	1	0	0	0	0	0
	TN	19	20	20	20	20	20
	TPR	0.00	0.10	0.10	0.00	0.10	0.00
	FPR	0.05	0.00	0.00	0.00	0.00	0.00

表 4.14: d = 10 の TPR と FPR

d=15							
		提案手法	方式2	方式3	方式5	方式8	方式9
始点	TP	5	7	7	6	8	6
	FN	15	13	13	14	12	14
	FP	1	0	0	0	0	0
	TN	19	20	20	20	20	20
	TPR	0.25	0.35	0.35	0.30	0.40	0.30
	FPR	0.05	0.00	0.00	0.00	0.00	0.00
終点	TP	1	3	3	0	5	1
	FN	19	17	17	20	16	19
	FP	1	0	0	0	0	0
	TN	19	20	20	20	20	20
	TPR	0.05	0.15	0.15	0.00	0.24	0.05
	FPR	0.05	0.00	0.00	0.00	0.00	0.00

表 4.15: d = 15 の TPR と FPR

d=20							
		提案手法	方式2	方式3	方式5	方式8	方式9
始点	TP	9	9	9	8	8	8
	FN	11	11	11	12	12	12
	FP	1	0	0	0	0	0
	TN	19	20	20	20	20	20
	TPR	0.45	0.45	0.45	0.40	0.40	0.40
	FPR	0.05	0.00	0.00	0.00	0.00	0.00
終点	TP	2	3	3	1	5	2
	FN	18	17	17	19	15	18
	FP	1	0	0	0	0	0
	TN	19	20	20	20	20	20
	TPR	0.10	0.15	0.15	0.05	0.25	0.10
	FPR	0.05	0.00	0.00	0.00	0.00	0.00

表 4.16: d = 20 の TPR と FPR

表4.11，表4.12，表4.13を見ると，提案手法では，1本のダンスシーンなしの音楽番組映像に対して，ダンスシーンの検出が発生してしまった．原因としては，検出されたシーンはバンドのパフォーマンスであり，楽器を演奏する際に，出演者の全身の動きが激しいため，発生した16×8MBは多くなり，ダンスシーンの16×8MBの特徴に似ていることであると考えられる．

表4.11，表4.12，表4.13における始点と終点のTPRとFPRに対して，FPRを横軸とし，TPRを縦軸とする散布図で表現した．各dに対して，始点の散布図は図4.3に示し，終点の散布図は図4.4に示す．

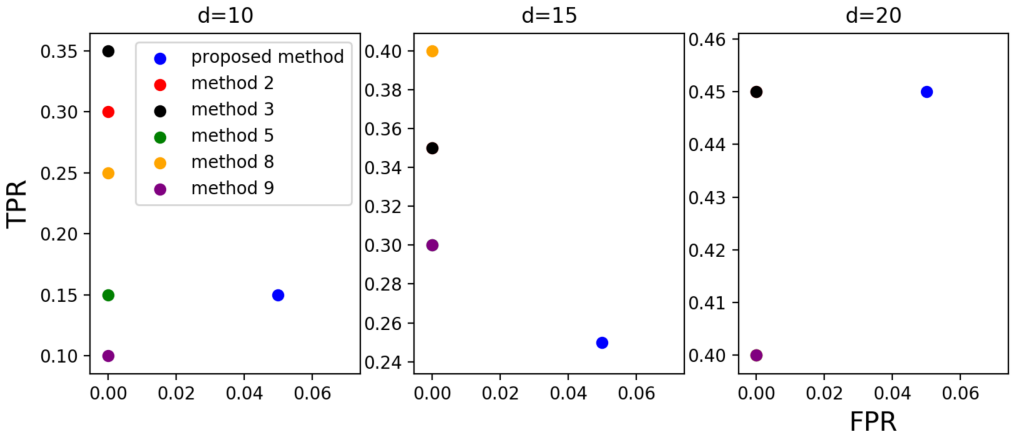

図 4.3: 始点の散布図

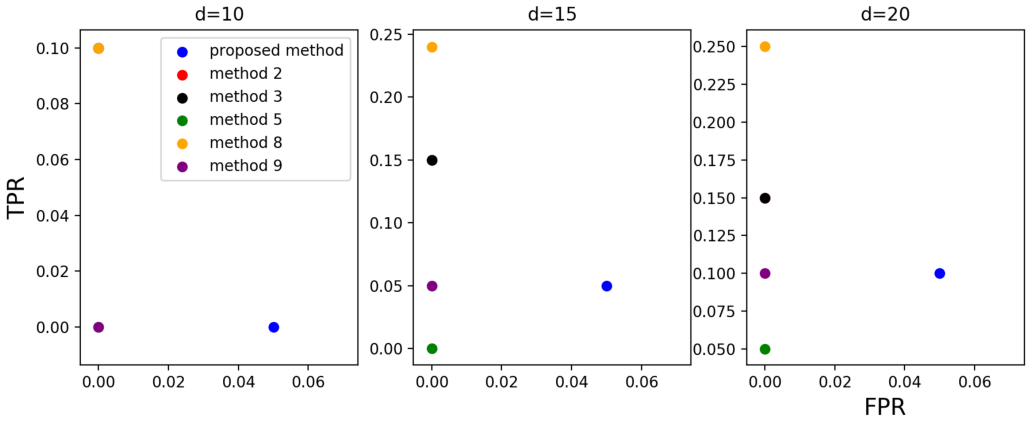

図 4.4: 終点の散布図

（FPR, TPR）は（0，1）に近づけば近づくほど，ダンスシーンを含まない音楽番組映像に対して誤検出は少なく，ダンスシーンを含む音楽番組映像に対して検出点が許容範囲に入っている映像は多いと言える．図 4.3 と図 4.4 を見ると，終点より始点の方はTPRが高い．原因としては，BSpline 曲線によるダンスシーンを検出する手法では，終点を検出する際に，該当の BSpline 曲線の極小値の平均を取っているため，精度が落ちると考えられる．また，d が大きくなると，始点と終点のTPRが高くなるが，許容範囲は広くなり，ユーザー体験が劣化する．d＝20 の場合，10 秒の許容幅であり，全ての方式に対して，始点のTPRは 0.40 以上である．とくに，方式 2 と方式 3 では，（FPR, TPR）は（0，0.45）であり，ほかの方式と比較すると，より優れる結果を得た．

第5章　特定ダンスシーンの検索

本章では，文献[3]の提案手法を用いて手作業で取り出した始点および終点の誤差のない複数のダンスシーンから検索キーと同一ダンスシーンの検索実験および結果について述べる．まず，5.1節で文献[3]の提案手法である3D重心曲線による動作動画の検索手法について述べる．5.2節で文献[3]の提案手法に基づいて複数のダンスシーンから検索キーと同一ダンスシーンを検索する手法と，実験およびその結果について述べる．

5.1　3D重心曲線による動作動画の検索

文献[3]では，3D重心曲線による動作動画の検索の流れは以下になっている．

Step1. 動画の各フレーム毎に$16\times8, 8\times16, 8\times8$画素のMBの出現座標と，それらのMB上で発生した動きベクトルの始点･終点座標を取得する．また，各MBの参照フレームがどのフレームであるかの情報も取得する．

Step2. 3Dモデルソフト[15]の三次元空間上において，x軸，z軸はフレーム画像内の座標とし，y軸は時間軸(フレーム番号)とすることで，動画各フレームの$16\times8, 8\times16, 8\times8$画素のMBは立方体のオブジェクト，動きベクトルは円柱のオブジェクトとして該当座標に設置する．

Step3. 1つのフレームに圧縮符号化パラメータのオブジェクト群の重心を求める作業を全てのフレームで行う．

Step4. 各フレームの重心を線形補間で繋げいた曲線を求める．

Step5. 検索キー動画の時系列重心座標データと，検索対象動画の時系列重心データ同士で相互相関関数を求めることにより，区間ごとに両データ間の距離を求める．

検索キー動画をf，検索対象動画をg，検索キー動画のフレーム数を

n，検索対象動画のフレーム数をN，スライドさせる数をiとすると，検索キー動画と検索対象動画の相互相関関数は以下のようになる．

$$F(i) = \sum_{t=0}^{n-1} \frac{(f_x(t)-g_x(t+i))^2+(f_y(t)-g_y(t+i))^2}{2} - n \times \overline{f} * \overline{g}_i \quad (i = 0, 1, ..., N-n) \quad (11)$$

ここで，

$$\overline{f} * \overline{g}_i = \frac{\left(\frac{\sum_{t=0}^{n-1} f_x(t)}{n} - \frac{\sum_{i=0}^{N-n} \frac{\sum_{t=0}^{n-1} g_x(t+i)}{n}}{N-n+1} \right)^2 + \left(\frac{\sum_{t=0}^{n-1} f_y(t)}{n} - \frac{\sum_{i=0}^{N-n} \frac{\sum_{t=0}^{n-1} g_y(t+i)}{n}}{N-n+1} \right)^2}{2} \quad (12)$$

5.2 ダンスシーン映像の3D重心曲線の観察

音楽TV番組におけるダンスシーン映像の先頭の6秒～17秒において，出演者が真正面から映り，カメラの動きが少ないという傾向にあるため，同一ダンスシーンである2つの映像の先頭シーンの3D重心曲線において，共通点があると考えられる．ダンスシーン映像の3D重心曲線を作る際に，舞台の背景とカメラワークにより発生した動きベクトルの影響を減らすため，5.1節Step1におけるMBは，固定した範囲内であるかつ一定サイズ以上の動きベクトルが発生したMBを限定した．図5.1に画面上に固定された範囲の例を示す.

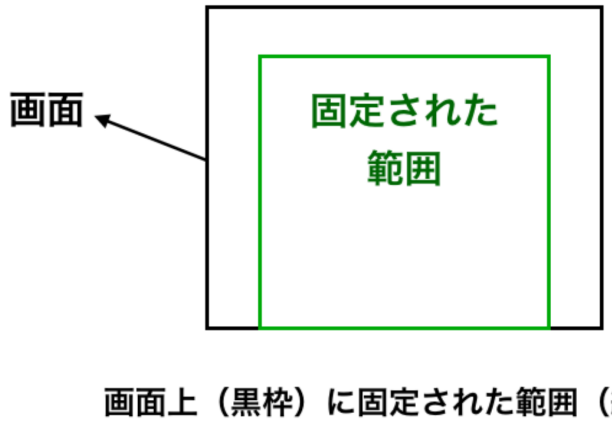

図 5.1: 画面（黒枠）上に固定された範囲（緑枠で囲まれた部分）の例

本節では，6本のダンスシーン映像を利用して，限定したMBとそれらのMB上で発生した動きベクトルの情報を用いて生成したダンスシーン映像の3D重心曲線を観察した．ここで，6本の映像は手作業で取り出した始点と終点の誤差のないダンスシーン映像であり，３つのペアからなる．その中に，映像1と映像2は同一ダンスシーン（振り付け）であり，映像3と映像4は同一ダンスシーンであるが，映像5と映像6は同一ダンスシーンである．それぞれのペアとする２つのダンスシーン映像に対して，振り付けは同一であるが，舞台の背景，照明，カメラワークと出演者の衣装それぞれは異なっている．6本のダンスシーン映像の情報は表5.1に示す．

ダンスシーン映像No.	1	2	3	4	5	6
フレーム数	4792	4995	5570	6980	5599	5657
曲	NO WAY MAN		U.S.A		your song	
出演者	AKB48		DA PUMP		嵐	
番組	うたコン	Mステ	紅白	Mステ	Mステ	Mステ

表 5.1: 6本のダンスシーン映像の情報

6本のダンスシーン映像の先頭の500フレームから生成した3D重心曲線の観察を行う．生成した3D重心曲線は図5.2，図5.3，図5.4，図5.5，図5.6，図5.7に示す．

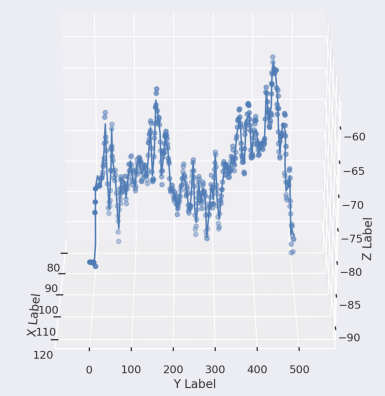

図 5.2: 映像1の3D重心曲線

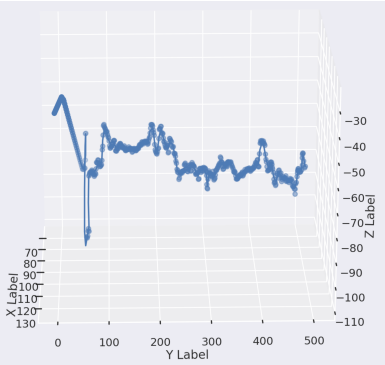

図 5.3: 映像2の3D重心曲線

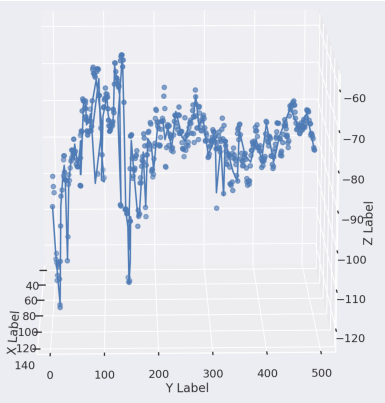

図 5.4: 映像3の3D重心曲線

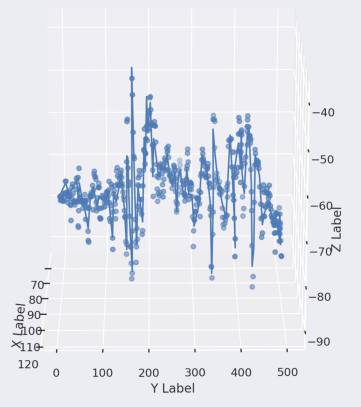

図 5.5: 映像4の3D重心曲線

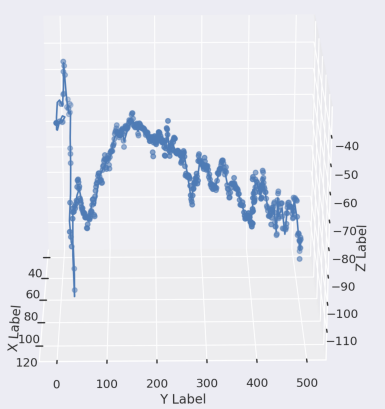

図 5.6: 映像 3 の 3D 重心曲線

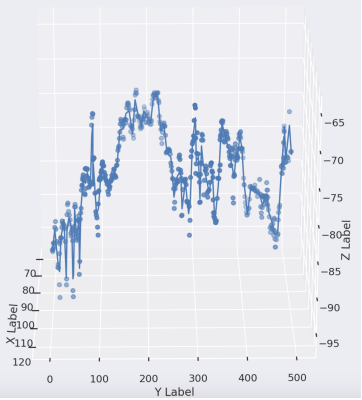

図 5.7: 映像 4 の 3D 重心曲線

図 5.2 と図 5.3 を見ると，y 座標は約 50 であるところに曲線が上昇し，約 200 のところに下向する共通点がある．図 5.3 と図 5.4 を見ると，y 座標は約 150 であるところに曲線が上昇し，約 320 のところに下向する共通点がある．図 5.5 と図 5.6 を見ると，図 5.5 において，y 座標は約 50 であるところに曲線が上昇し，約 280 のところに下向している．図 5.6 において，y 座標は約 100 であるところに曲線が上昇し，約 280 のところに下向している．

5.3 特定ダンスシーン検索の手順

特定ダンスシーンを検索する際，検索キーについては，既存のダンスシーンの先頭の n 枚フレームからなる動画クリップを検索キーとする．

文献 [3] の提案手法に基づいた特定ダンスシーンの検索手順は以下になっている．特定ダンスシーンの検索手順において，MB_x は MB の x 座標を表し，MB_y は MB の y 座標を表す．

Step1. 動画の各フレーム毎に，$15 < MB_x < 75 \wedge 10 < MB_y < 70$ の範囲内であることかつ MB 上で発生した動きベクトルのサイズは 2 以上であることを満たす 16×8MB の出現座標と，それらの MB 上で発生した動きベクトルの始点・終点座標を取得する．また，各 MB の参照フレームがどのフレームであるかの情報も取得する．

Step2. 3D モデルソフトの三次元空間上において，x 軸，z 軸はフレーム画像内の座標とし，y 軸は時間軸 (フレーム番号) とすることで，動画各フレームの 16×8MB は立方体のオブジェクト，動きベクトルは円柱のオブジェクトとして該当座標に設置する．

Step3. 1つのフレームに圧縮符号化パラメータのオブジェクト群の重心を求める作業を全てのフレームで行う.

Step4. 各フレームの重心を線形補間で繋げいた曲線を求める.

Step5. 各検索キー動画の時系列重心座標データと，ほかの5本の検索対象動画の時系列重心データ同士で相互相関関数を求めることにより，区間ごとに両データ間の距離を求める.

検索キー動画をf，検索対象動画をg，検索キー動画のフレーム数をn，検索対象動画のフレーム数をN，スライドさせる数をn回とすると，式（11）は以下のようになる.

$$F(i) = \sum_{t=0}^{n-1} \frac{(f_x(t)-g_x(t+i))^2+(f_y(t)-g_y(t+i))^2}{2} - n \times \overline{f} * \overline{g}_i \text{(i=0,1,...,n-1)} \qquad (13)$$

ここで，

$$\overline{f} * \overline{g}_i = \frac{\left(\frac{\sum_{t=0}^{n-1} f_x(t)}{n} - \frac{\sum_{i=0}^{n-1} \frac{\sum_{t=0}^{n-1} g_x(t+i)}{n}}{n}\right)^2 + \left(\frac{\sum_{t=0}^{n-1} f_y(t)}{n} - \frac{\sum_{i=0}^{n-1} \frac{\sum_{t=0}^{n-1} g_y(t+i)}{n}}{n}\right)^2}{2} \qquad (14)$$

Step6. 1つの検索キーに対して，式（13）によって得た5つの$F(i)$（$i = 0, 1, 2, ...n$）の結果の中から，最も小さい最小値をもつ$F(i)$に対応するダンスシーン映像は検索キーと同一ダンスシーンであることを判定する.

5.4 特定ダンスシーン検索の実験および考察

本研究では，5.3節で述べた手法を用いて，特定ダンスシーンの検索を行う．使用した映像は5.2節で使用した6本のダンスシーン映像である．検索キー1~6それぞれは，ダンスシーン映像1~6の前n枚のフレームからなる動画クリップに対応する．各検索キーに対して，式（13）によってほかの5本のダンスシーン映像との相互相関関数を求めた．検索キーのフレーム数nを100，150，200，250，300，400，500にして行った実験の$F(i)$（$i = 0, 1, 2, ...n$）の結果それぞれは図5.8，図5.9，図5.10，図5.11，図5.12，図5.13，図5.14に示す．

各図において，(a)~(f)は検索キー1~6とほかの5本のダンスシーン映像の相互相関関数の結果を示す．各(a)~(f)において，x軸はiであり，y軸は$F(i)$であり，各色線は該当検索キーとほかの5本ダンスシーン映像の相互相関関数の折れ線である．各色線に対応するダンスシーン映像番号の情報は表5.2に示す．

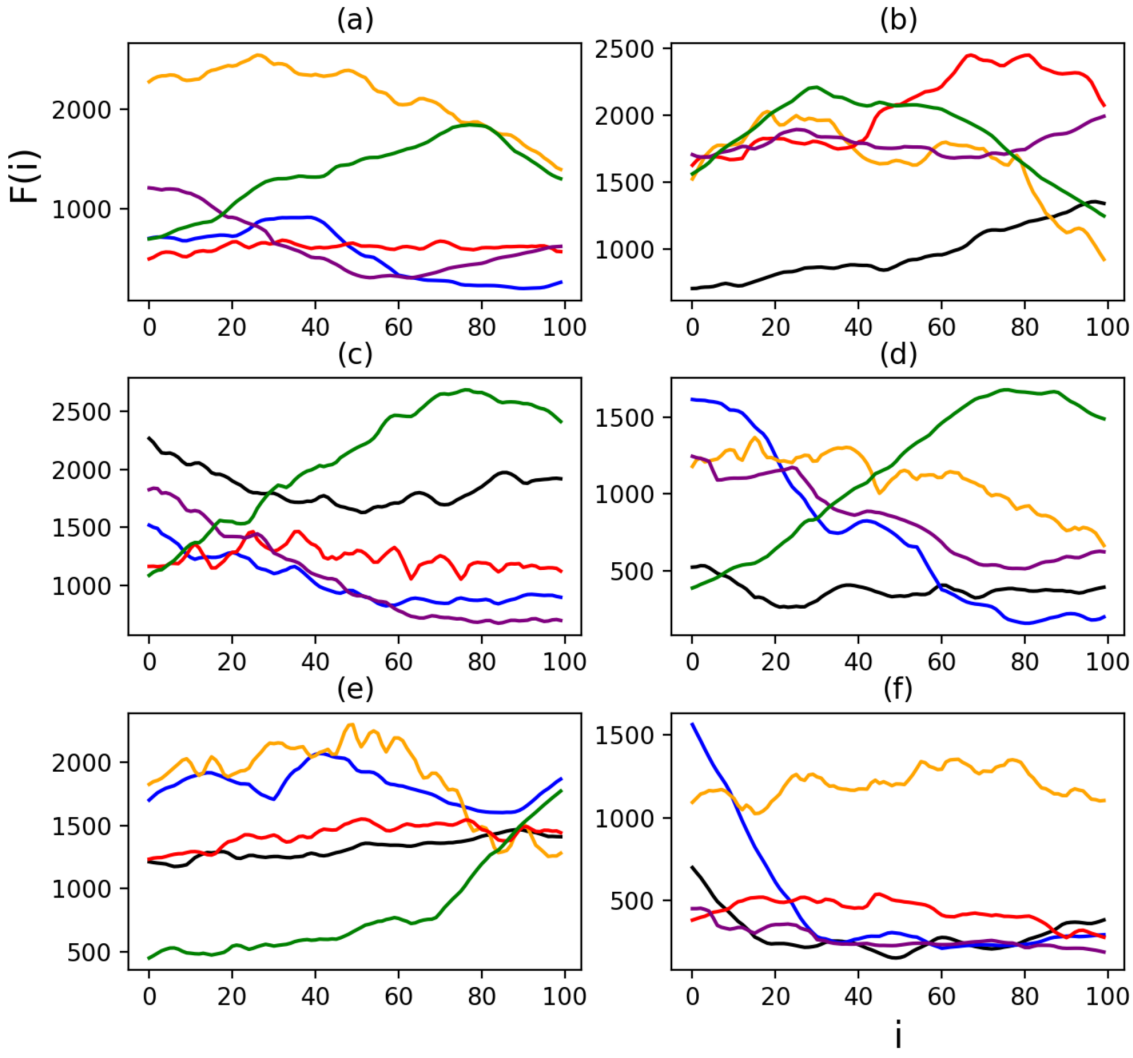

図 5.8: $n = 100$ の $F(i)$

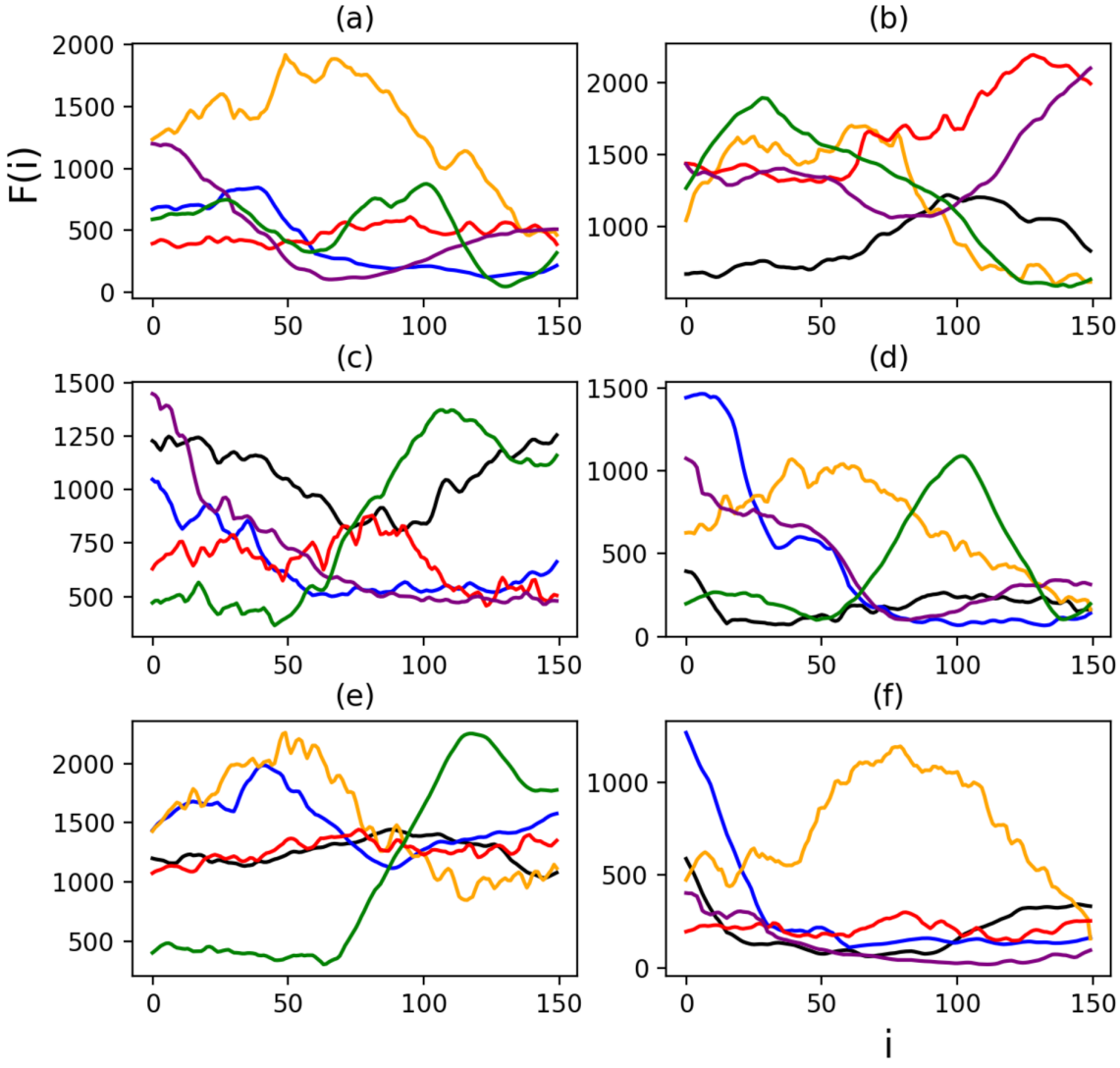

図 5.9: $n = 150$ の $F(i)$

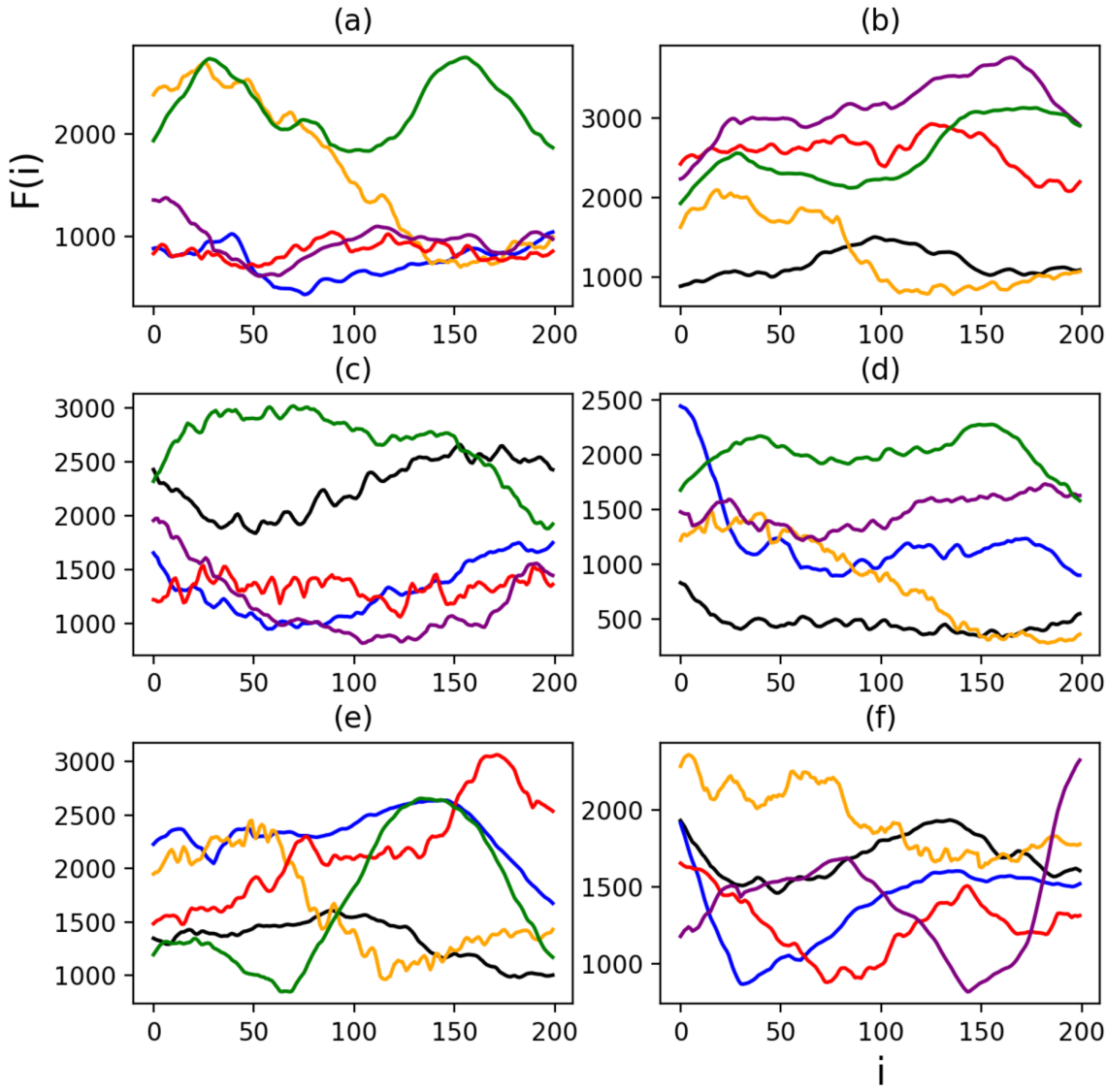

図 5.10: $n = 200$ の $F(i)$

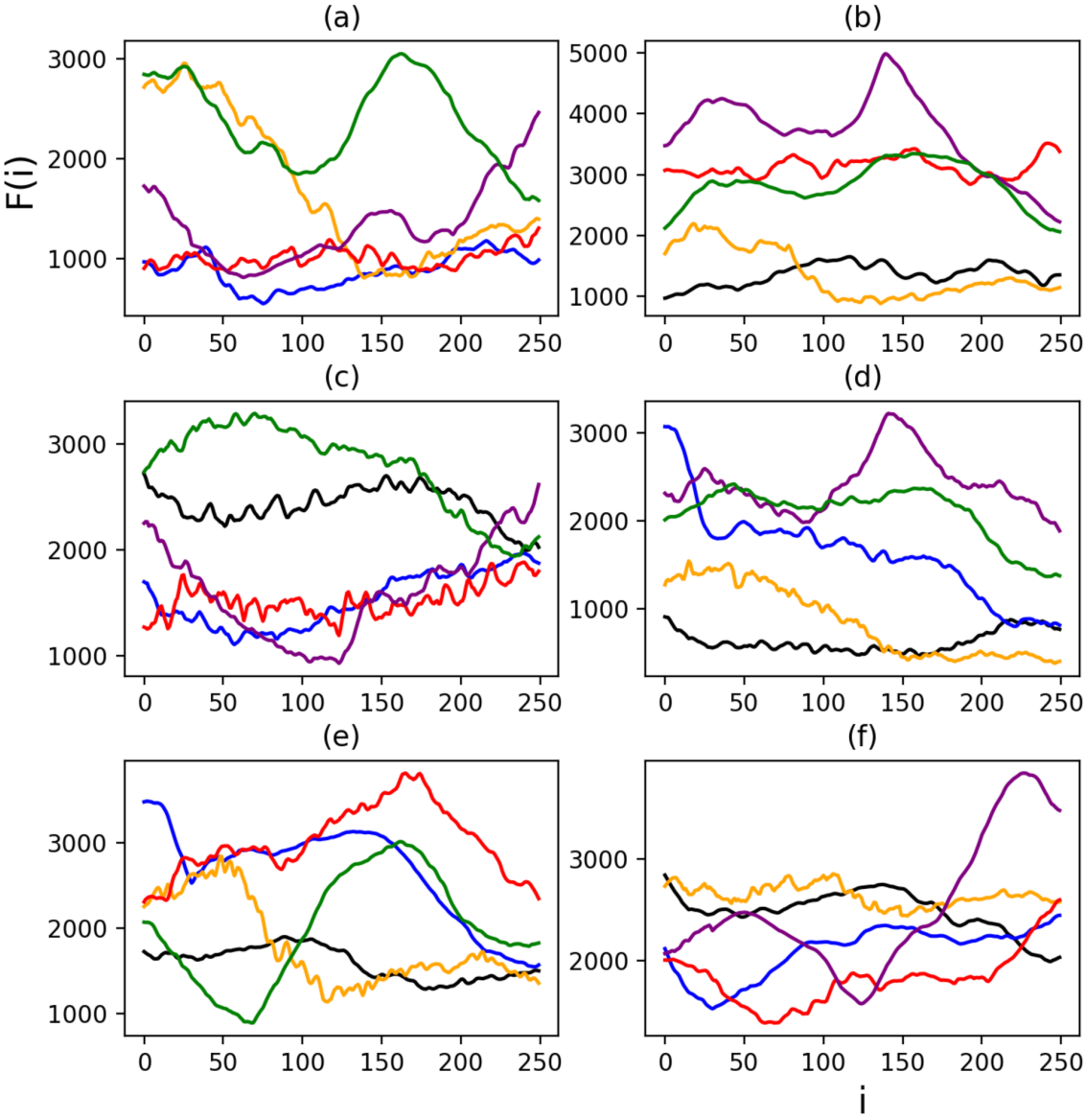

図 5.11: $n = 250$ の $F(i)$

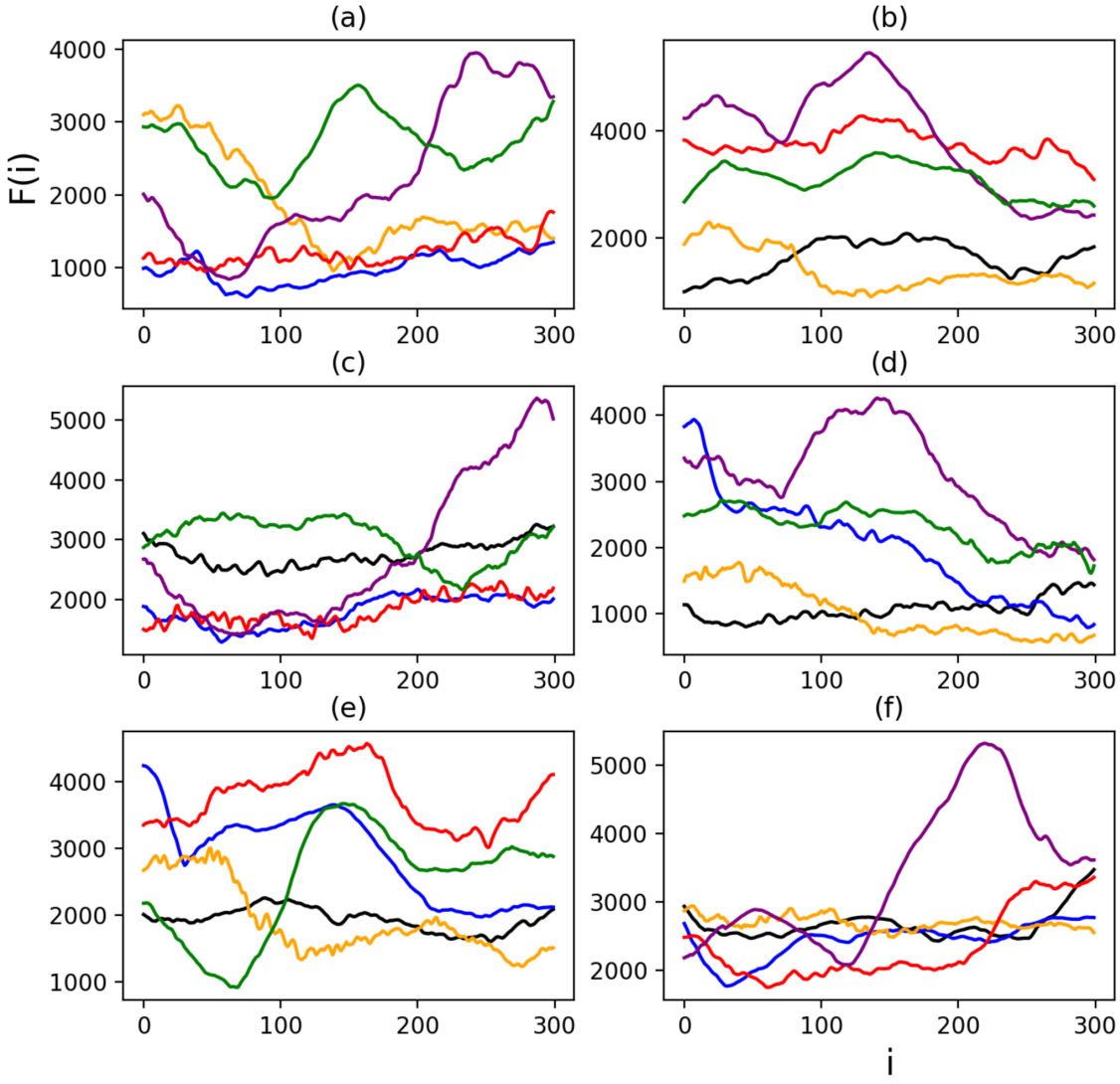

図 5.12: $n = 300$ の $F(i)$

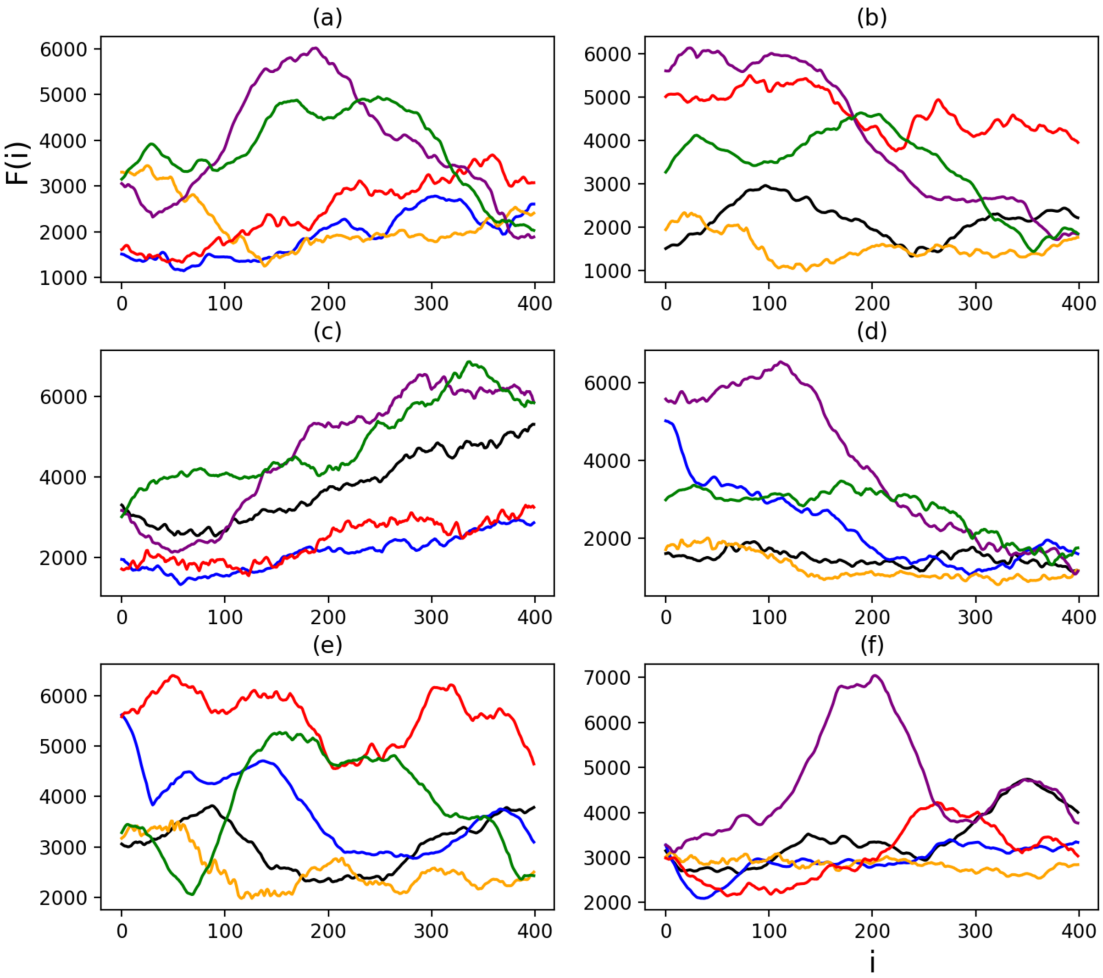

図 5.13: $n = 400$ の $F(i)$

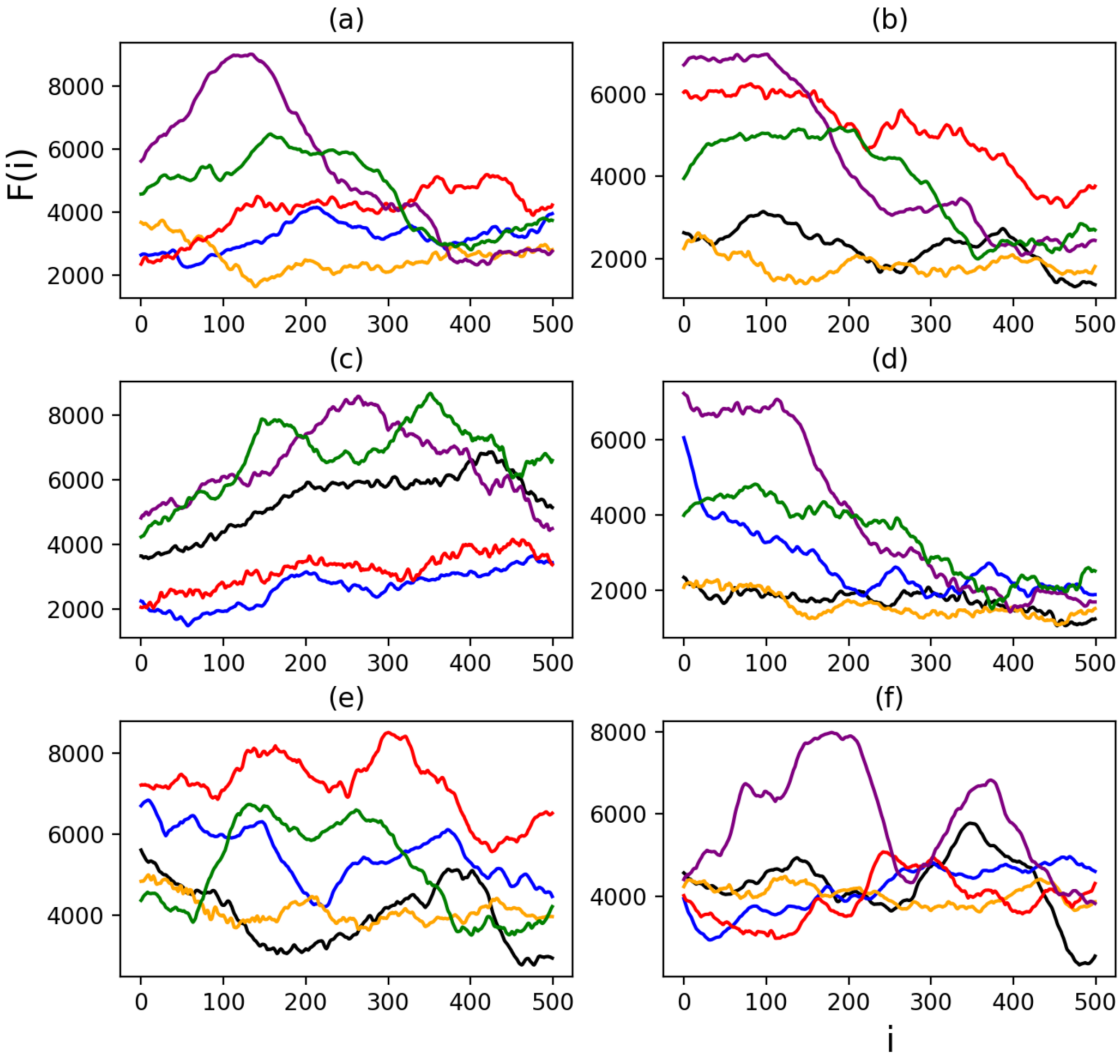

図 5.14: $n = 500$ の $F(i)$

線	黒線	青線	オレンジ線	赤線	紫線	緑線
対応するダンスシーン映像の番号	1	2	3	4	5	6

表 5.2: 色線の情報

各nに対して，$F(i)$の最小値による各検索キーと同一ダンスシーンであると判断された映像の番号は表5.3，表5.4に示す．表5.3と表5.4における1～6の番号は6本のダンスシーン映像の番号に対応する．赤数字は正しく検索できたことを表す．

	n=100	n=150	n=200	n=250
検索キー1	2	6	2	2
検索キー2	1	3	3	3
検索キー3	5	6	5	5
検索キー4	2	2	3	3
検索キー5	6	6	6	6
検索キー6	1	5	5	4

表 5.3: $F(i)$の最小値による検索された結果I

	n=300	n=400	n=500
検索キー1	2	2	3
検索キー2	3	3	1
検索キー3	2	2	2
検索キー4	3	3	1
検索キー5	6	3	1
検索キー6	4	2	1

表 5.4: $F(i)$ の最小値による検索された結果 II

表5.3と表5.4を見ると，$n = 200$ の場合，検索キー1とダンスシーン映像2，検索キー4とダンスシーン映像3，検索キー5とダンスシーン映像6，検索キー6とダンスシーン映像5が同一ダンスシーンであることを判定できた．一方，検索キー2とダンスシーン映像3，検索キー3とダンスシーン映像5が同一ダンスシーンであることを判定してしまい，誤判であった．$n = 200$ の場合がより多くの正しい結果が得られた一方で，$n = 400$ の場合は 2 つの正しい結果，$n = 500$ の場合は1つだけの正しい結果が得られた．それはnが大きくなると，カメラワークを使用した回数も増えて，出演者の全身の動き情報を取りにくくなるためであると考えられる．また，nが小さくなると，出演者の動き情報が少ないため，判定するのは困難であると考えられる．

第6章　まとめ

6.1　まとめ

本研究では，音楽TV番組における特定ダンスシーンの検索システムを構築するには，ダンスシーンの検出と特定ダンスシーンの検索実験を行い，評価をした．

圧縮動画におけるダンスシーンでは，予測16×8MBが発生しやすく，ダンスシーンの始まりと終わりのGOP番号において，急激な16×8MB数の変化が発生する傾向があるため，ダンスシーンの検出実験では，平滑化z-スコアアルゴリズムの移動平均部分と，BSpline曲線を組み合わせることによってダンスシーンの検出をできた．4.4節での提案手法であるℓ=30，w=0.8，N_B=50の方法を用いた実験結果のF値の平均が0.83となる．提案手法の改善方法について，提案手法を用いて得たダンスシーン検索の結果を基準にして，改善方法の誤検出を減らした上で，移動平均の50個の点列を用いて生成したBSpline曲線上における等間隔に選んだ1000個の出力点を利用してダンスシーンの検出手法では，F値の平均が0.88となり，より良い結果が得られた．

特定ダンスシーンの検索実験では，文献[3]で提案手法に基づき，3D重心曲線より特定ダンスシーンの検索を行った．6本のダンスシーン映像を検索対象として，それぞれの映像の前200フレームからなる動画クリップを検索キーとする場合，その中の4つの検索キーに対して，同一ダンスシーンである映像を検索できた．

本研究を通して，圧縮動画の符号化パラメータを特徴量とし，復号化をせずに音楽ライブ番組からダンスシーンの検出が可能であることがわかった．また，特定ダンスシーンの検索については，検索キーとする動画クリップの区間および長さを変えることによって，検索性能を上回ることが期待できる．

6.2 今後の課題

ダンスシーンの検出実験では，入力点列の個数と出力点列の点数を変更するとより良い結果を得られることが分かったが，BSpline曲線の微分曲線を用いてさらに精度を上げることができると考えられるため，今後，BSpline曲線の微分曲線を作成し，微分曲線によるダンスシーンの検出を検討していきたい．

特定ダンスシーンの検索実験では，使用した映像は手作業で取り出した始点と終点の誤差のないダンスシーン映像であるため，ダンスシーン検出実験で検出された映像のマッチングに対応できない．1つのダンスシーンでは，特徴的な振り付けが出る際に，カメラが静止している可能性がかなり大きい．そこで，カメラワークを解析することで，特徴的な振り付け区間を発見し，ダンスシーン検出実験で検出された映像とのマッチングを取ることが今後の課題である．

謝辞

本研究の実施および本論文の作成にあたり，研究の進む方向，資料の作成，発表の仕方，論文の書き方など熱心で適切なご指導，ご助言を下さった森田啓義先生に心より感謝致します. 留学生である私の質問に対して，いつもわかりやすく説明していただき，心より感激致します.

また，ゼミ等で発表の仕方などご助言を賜りました笠井裕之先生，研究室内の様々なお世話をしてくださった事務の片桐さんに深く御礼申し上げます.

最後に，多くのアドバイス，励まし言葉をいただいた森田研究室，笠井研究室の学生の皆様，先行研究を参考にさせて頂いた先輩・研究者の皆様に深く感謝致します.

参考文献

[1] 保坂理人,“MPEG2圧縮された高画質サッカー映像からのハイライトシーン検出”,電気通信大学大学院情報システム学研究科修士論文(2012).

[2] 祖泉大河,“時空間マクロブロックパターンを用いたシーン検出”,電気通信大学情報・通信工学科卒業論文(2018).

[3] 岡村和磨,“人の動作ビデオクリップを検索キーとする圧縮動画検索”,電気通信大学大学院情報理工学研究科修士論文(2018).

[4] 単鴻,“MPEG2動きベクトルを用いた複数移動体の検知・追跡システム”,電気通信大学大学院情報システム学研究科修士論文(2008).

[5] 吉田壮, 小川貴弘, 長谷山美紀,“歌謡番組における映像の構造に注目したシーン分割手法 ”,電子情報通信学会論文誌，Vol.J97-D , No.7 (2014).

[6] 平井辰典, 中野倫靖, 後藤真孝，森島繁生,“歌手映像と歌声の解析に基づく音楽動画中の歌唱シーン検出手法の検討”, 情報処理学会研究報告 , Vol.n, No.n (2014).

[7] 瀬上隆博，グラウンド境界線の傾斜度を用いたMPEG2サッカー映像からのイベントシーン検出, 電気通信大学大学院情報システム学研究科修士論文(2010).

[8] MPlayer, http://www.mplayerhq.hu/

[9] FFMPEG, http://ffmpeg.mplayerhq.hu/

[10] Patrick Perkins, Steffen Heber,“Identification of Ribosome Pause Sites Using a Z- Score Based Peak Detection Algorithm”, 2018 IEEE 8th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) (2018).

[11] Mehmet Cem Catalbas, Tomaz Cegovnik, Jaka Sodnik, Arif Gulten, "Driver fatigue detection based on saccadic eye movements", 2017 10th International Conference on Electrical and Electronics Engineering(ELECO) (2017)

[12] SciPy. org,
https://docs.scipy.org/doc/scipy/reference/generated/
scipy.interpolate.BSpline.html

[13] Tajima Robotics,
https://tajimarobotics.com/basis-spline-interpolation/ (2018/7/27).
https://tajimarobotics.com/basis-spline-interpolation-2/ (2018/7/29).
https://tajimarobotics.com/basis-spline-interpolation-example/ (2018/8/5).
https://tajimarobotics.com/basis-spline-interpolation-program/ (2018/8/5).
https://tajimarobotics.com/basis-spline-interpolation-program-2/
(2018/12/8).

[14] Tom Lyche and Knut Morken, Spline methods,
http://www.uio.no/studier/emner/matnat/ifi/INF-MAT5340/v05/
undervisningsmateriale/

[15] blender, https://www.blender.org